TOPICS IN MULTIRESPONSE ANALYSIS AND OPTIMIZATION

By

ELSIE S VALEROSO

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1995

UNIVERSITY OF FLORIDA LIBRARIES

© Copyright 1996

by

Elsie S. Valeroso

ACKNOWLEDGEMENTS

I am deeply grateful to the chairman of my supervisory committee Dr André I Khuri, for his guidance, patience and support throughout the writing of this dissertation. I thank the other members of my committee, Dr John Cornell Dr Geoffrey Vining, and Dr Mang Tia for their patience and assistance, and also Professor Victor Chew who has been very helpful to me. I am grateful to the following individuals for their aid and kindness Dr Ron Randles Ms Fe Lorica, Dr Larry Winner, Ms Maria Ripoll, Dr Jim Hobert, Professor Paul Chen of Oregon State University, Mr Bill Miller of the USDA/ARS Laboratory in Orlando, Dr Deonis Mermo of the Southeastern Louisiana University, and Dr John Borkowski of Montana State University Finally I thank my parents, siblings, and Dr Emmanuel Bugtan for their love support and encouragement

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

ABSTRACT vii

CHAPTERS

1 INTRODUCTION 1

1.1 An Overview of the Present Work 1
1.2 Objectives of the Present Work 5

2 REVIEW OF LITERATURE 6

2.1 A Note on Response Surface Methodology 6
2.2 Multiresponse Experiments 8
2.3 Multiresponse Parameter Estimation 10
2.4 Multiresponse Optimization 11
2.5 Blocking in RSM 18
2.6 Multiresponse Design Criteria 20

3 OPTIMIZATION OF MULTIRESPONSE FUNCTIONS 24

3.1 Introduction 24
3.2 The Multiresponse Model 25
3.3 Procedures for the Simultaneous Optimization of the Responses 28
3.4 Numerical Examples 31

4 MULTIRESPONSE MODELS WITH FIXED BLOCK EFFECTS 48

4.1 Introduction 48
4.2 Estimating the Parameters in Multiresponse Models with Fixed Block Effects 49
4.3 Estimating the Mean Response in the Presence of Fixed Block Effects 52
4.4 Effect of Blocking on the Prediction Variance 54
4.5 Orthogonal Blocking 56
4.6 A Numerical Example 60

5 MULTIRESPONSE MODELS WITH RANDOM BLOCK EFFECTS 66

5.1 Introduction 66
5.2 Estimating the Parameters in Multiresponse Models with Random Block Effects 67
5.3 Effect of Orthogonal Blocking on the Estimation of Parameters 69
5.4 Estimating the Parameters in Multiresponse Models with Mixed Effects 78
5.5 Mean Response and Prediction Variance 84
5.6 Optimization 85
5.7 Closeness to Target Values 86
5.8 A Numerical Example 86

6 DESIGN CRITERION TO DECREASE THE INTEGRATED MEAN SQUARED ERROR 91

6.1 Introduction 91
6.2 Model and Notation 92
6.3 Integrated Mean-Squared Error Criterion 93
6.4 Minimization of J 97
6.5 A Numerical Example 99
6.6 Other Possible Approaches for the Minimization of J 104

7 CONCLUSIONS AND FUTURE RESEARCH 107

7.1 Conclusions 107
7.2 Future Research 108

A MAXIMUM LIKELIHOOOD ESTIMATOR OF Σ 110

B SOME RESULTS FOR MULTIRESPONSE MODELS 112

B.1 PRESS and Cp criteria for Multiresponse models 112
B.2 McElroy's R^2 116
B.3 Optimization Algorithm by Price 118

C DESIGN EXPERT RESULTS FOR CHAPTER 3, SECTION 3.4 122

D PROOFS OF THEOREMS 4.5.1 AND 4.5.2 129

D.1 Proof of Theorem 4.5.1 129
D.2 Proof of Theorem 4.5.2 130

E SAS RESULTS FOR CHAPTER 4, SECTION 4.6 137

F ESTIMATES OF VARIANCE-COVARIANCE MATRICES FOR SECTION 5.8 141

REFERENCES 143

BIOGRAPHICAL SKETCH 147

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

TOPICS IN MULTIRESPONSE ANALYSIS AND OPTIMIZATION

By

Elsie S Valeroso

August 1996

Chairman André I Khuri
Major department Statistics

Four topics concerning statistical techniques for the analysis of multiresponse experiments are presented. These techniques can be used to (1) set up a series of multiresponse experiments (design) that will yield adequate and reliable predictions of the responses of interest on the basis of a certain hypothesized model (2) fit the hypothesized multiresponse model using data collected according to the design chosen in (1), and (3) determine optimum operating conditions on the input variables that lead to the simultaneous optima of the predicted values of the responses, within a certain region of interest. The first topic is an extension of the optimization procedure discussed by Khuri and Conlon (1981). The simultaneous optimization of multiresponse functions represented by polynomial regression models that are not necessarily of the same form and degree are considered. The optimization procedure involves the minimization of a distance function that measures the distance between the vector of estimated responses and the assumed 'ideal' optimum over the experimental region. The second topic deals with multiresponse models with block effects that are either fixed or random. The effect of blocking on the estimation of the mean response, the

prediction variance, and on the simultaneous optima of the estimated multiresponse function is discussed The final topic is the construction of designs to minimize a multiresponse version of the integrated mean squared error This is an extension of the work done by Kim and Draper (1994) to cases where more than two responses and more than one input (control) variable of interest are considered, and also where the variance-covariance matrix associated with the responses is unknown A sequential procedure wherein the design points are chosen iteratively is presented

CHAPTER 1

INTRODUCTION

1.1 An Overview of the Present Work

Much has been written concerning response surface methodology (RSM) involving a single response. The review article by Myers et al. (1989) shows how much work has been done in this particular area. However, the same cannot be said about RSM involving several responses. The design and analysis of experiments with multiple responses has received very little attention in the statistical literature. Yet, in many experiments it is not uncommon to measure more than one response from the same experimental unit. These experiments are called *multiresponse experiments*. For example, in a chemical process, there are almost always several properties of the product that are of interest such as the yield and the cost of making the product. In the development of food products the food technologist might consider in addition to the flavor of the product, the shelf life and subsequent undesirable byproducts. In the pharmaceutical area, a scientist is not only concerned about the efficacy of the drug but also of the possibility of serious side effects on the patient. These are but a few of the situations where multiresponse experiments are used. With the recent advances in technology, obtaining data that describe the different facets of a system or product is becoming more commonplace. Thus it is very important that multiresponse techniques be developed and made available to researchers and data analysts.

It is therefore the purpose of this dissertation to introduce new multiresponse techniques which presently are missing in the literature on the design and analysis of multiresponse experiments

When several responses are considered simultaneously, they should not be treated as independent, particularly when it is suspected they are correlated In multiresponse optimization, for example, in order to determine optimum operating conditions, it is not sufficient to only determine individual optima of the responses This is because optimal conditions for one response may be far from optimal or even physically impractical for the other responses Also, in the design and analysis of multiresponse experiments, a design that is satisfactory for one response may produce unsatisfactory results for the rest of the responses. Thus any design criterion used should be based on the consideration of all the responses together rather than individually

One important objective of multiresponse experimentation is the determination of optimum operating conditions on the input (control) variables that will lead to the simultaneous optimization of the predicted values of the responses within a certain region of interest. For example, in a tool life problem, the experimenter may wish to determine which cutting speed and depth of cut will maximize the lifetime of the tool and also at the same time maximize the rate of material removed Another example is the chemical process situation described above The scientist may want to find the levels of temperature and pressure which both maximize the yield and minimize the cost of the process. There are many examples of practical situations where multiple responses are observed and where there is a need to reach some type of compromise as far as optimum conditions are concerned Several approaches are available in the statistics literature which can be used to determine these optimal conditions These approaches include the ones given by Myers and Carter (1973), Derringer and Suich (1980), Khuri and Conlon (1981) and more recently, the one by Del Castillo and Montgomery (1993) A review of these procedures is given in Chapter 2 In Chapter

3, another multiresponse optimization procedure is introduced This procedure is an extension of the work done by Khuri and Conlon (1981) to the case where the response models proposed are not of the same form and degree.

There are many experimental situations where it is impossible to perform all the runs under homogeneous conditions Examples of homogeneous conditions might be a single period of time, say one day, a single batch of homogeneous raw material, or using a single machine or operator in a particular process. Multiple days or batches of raw material represent possible extraneous sources of variation (day, batch, etc) and are called block effects. Such blocking conditions need to be considered in the analysis of the multiresponse experiment Blocking may also be used in instances where it may be better to deliberately vary the experimental conditions to ensure that the treatments are equally effective across many situations that may be encountered in practice For example, a chemical engineer may run a pilot plant experiment with several batches of raw material because he/she knows that different batches of raw material will be used in the full scale process However, very little information is available on the analysis of multiresponse experiments with block effects Results using blocking in single-response experiments are given by Box and Hunter (1957, 1961a,b) and Khuri (1992,1994,1996b) but nothing much has been done about the multiresponse case. The importance of considering block effects in single-response experiments has long been established and the same should be done for multiresponse experiments. A review of blocking in RSM is given in Chapter 2. Chapters 4 and 5 deal with the analysis of multiresponse models with block effects The results given in these two chapters are the multiresponse generalizations of the works done by Khuri (1992 1994, 1996b)

Prior to running a multiresponse experiment, the researcher must first decide which experimental design to use The task of selecting a suitable response surface design is not an easy one There are many classes of experimental designs available in

the literature. As a result, Myers and Montgomery (1995, p 280) offer some guidelines on the choice of a response surface design (see also Box and Draper (1975)). They list the following as important characteristics of response surface designs:

1. Result in a good fit of the model to the data
2. Give sufficient information to allow a test for lack of fit
3. Allow models of increasing order to be constructed sequentially
4. Provide an estimate of pure experimental error
5. Be insensitive (robust) to the presence of outliers in the data.
6. Be robust to errors in control of design levels.
7. Be cost effective
8. Allow for experiments to be done in blocks
9. Provide a check on the homogeneous variance assumption.
10. Provide a good distribution of the variance of the predicted values.

There are also other criteria on which the choice of designs is based. For example, the researcher may use the D, A, E, G and J design optimality criteria. However, most of the results in these design optimality criteria apply only to the single-response case. A few results are also available for multiresponse experiments. Wijesinha and Khuri (1987a) presented results concerning multiresponse D-optimal designs and Kim and Draper (1994) worked on the J optimality criterion. A description of the results obtained by Kim and Draper (1994) and a generalization of their work to other cases are given in Chapter 6.

1.2 Objectives of the Present Work

The main purpose of this research is to develop new techniques that can be used to

1. set up a series of multiresponse experiments (design) that will yield adequate and reliable predictions of the responses of interest on the basis of a certain hypothesized model;

2. fit the hypothesized multiresponse model using data collected according to the design chosen in (1), and

3. determine optimum operating conditions on the input variables of the model that result in the simultaneous optimization of the predicted values of the responses, within a certain region of interest.

The three objectives listed above are addressed in Chapters 3 to 6 of this dissertation. The problem of setting up designs, stated in objective 1 is tackled in Chapter 6. A sequential procedure for seeking a design that minimizes the integrated mean-squared error is presented in Chapter 6. The second objective is addressed in Chapters 4 and 5. These two chapters deal with the fitting of multiresponse models that take into account extraneous sources of variation, i.e., block effects. The results given are the multiresponse generalizations of the works done by Khuri (1992, 1994, 1996b) in the single-response case. In Chapter 3, a multiresponse optimization procedure is introduced to obtain the simultaneous optima of the responses of interest. This procedure is an extension of the work done by Khuri and Conlon (1981) to the case where the multiresponse models used are not of the same form and degree.

CHAPTER 2

REVIEW OF LITERATURE

2.1 A Note on Response Surface Methodology

Response surface methodology (RSM), as defined by Myers and Montgomery (1995, p 1), is a collection of statistical techniques useful for developing, improving, and optimizing processes. These techniques, as described by Khuri and Cornell (1996, p 3), encompass

1. setting up a series of experiments (designing a set of experiments) that will yield adequate and reliable measurements of the response of interest
2. determining a mathematical model that best fits the data collected from the design chosen in (1), and
3. determining the optimal settings of the experimental factors that produce the maximum or minimum value of the response

The formal development of RSM started with the work of Box and Wilson (1951). However, its origin can be traced back to the works of J. Wishart, C. P. Winsor, F. Yates, and others in the early 1930s. RSM was initially developed for the purpose of determining optimum operating conditions in chemical industry applications. Now it is used in a variety of fields and applications, not only in the physical and engineering sciences, but also in the biological, clinical, and social sciences.

Applications of RSM are mostly seen in situations where several input variables potentially influence some performance measure or characteristic of a product or process, that is, input variables influencing one or more response variables. Responses are usually measured on a continuous scale. The input variables, also called control or independent variables, are subject to the control of the experimenter.

The basic response surface problem involves a response of interest, y^*, which is a function of a vector $\boldsymbol{\zeta}^* = (\zeta_1^*, \zeta_2^*, \ldots, \zeta_k^*)$ of k input variables. The relationship is given by $y^* = f^*(\boldsymbol{\zeta}^*) + \epsilon^*$ where the form of f^* is unknown and ϵ^* is a term that represents other sources of variability not accounted for in f^*. It is usually the case where ϵ^* is treated as random error so that it is assumed that it follows the normal distribution with mean zero and variance σ^2. Thus $E(y^*) = \eta^* = f^*(\boldsymbol{\zeta}^*)$. The input variables $\zeta_1^*, \zeta_2^*, \ldots, \zeta_k^*$ are said to be in their natural form if they are expressed in terms of the natural units of measurement, such as degrees Celcius or pounds per square inch. It is convenient to analyze data without dealing with different units of measurement of the input variables. In this case, $\zeta_1^*, \zeta_2^*, \ldots, \zeta_k^*$ are transformed to their coded form $x_1, x_2, \ldots, x_k$. These coded variables are scale free, are centered at the origin of the coordinate axes, and have equal spreads in all directions of the coordinate axes. In terms of the coded variables, the true response function can be written as $\eta^* = f^*(\boldsymbol{x})$ where $\boldsymbol{x} = (x_1, x_2, \ldots, x_k)$.

Because the true form of η^* is unknown, it is assumed that η^* can be approximated by a polynomial model in k variables. Polynomial models derive their form by recognizing that any continuous, smooth function f^* for which $\eta = f^*(x_1, x_2, \ldots, x_k)$ can be *approximated locally* about a point in $x_1, x_2, \ldots, x_k$ space with a Taylor series expansion of terms involving $x_1, x_2, \ldots$ and x_k.

Generally speaking, a term in the polynomial is said to be of degree d if it contains the product of d of the factors $x_1, x_2, \ldots, x_k$ or if d is equal to the sum of the powers of the factors in the term. For example, with three factors x_1, x_2, and x_3, terms of the degree

2 are x_1^2, x_2^2, x_3^2, x_1x_2, x_1x_3, and x_2x_3, whereas terms of degree 3 are x_1^3, x_2^3, x_3^3, $x_1x_2x_3$, $x_1^2x_2$, $x_1^2x_3$, etc. A polynomial is said to be of degree d if the model contains terms up to degree d. Hereafter, when we refer to a polynomial model of degree d, we shall mean a model that contains *all terms in the factors up to and including terms of degree d*. Polynomial models of degree 1, 2, and 3 in the two variables x_1 and x_2 are

$$\eta^* = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$\eta^* = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2$$

$$\eta^* = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \beta_{111} x_1^3 + \beta_{222} x_2^3 + \beta_{112} x_1^2 x_2 + \beta_{122} x_1 x_2^2$$

respectively, where β_0, β_j, β_{jj}, β_{jjj}, β_{jl}, and β_{jjl}, $j, l = 1, 2$ are unknown parameters or model coefficients to be estimated from the data. We also note that a *reduced form* of a polynomial of degree d is a model containing only some or a subset of the terms of a polynomial of degree d.

The response surface model may be written in the form

$$\eta^* = f^*(\boldsymbol{x}, \boldsymbol{\beta}^*)$$

or equivalently,

$$y^* = f^*(\boldsymbol{x}, \boldsymbol{\beta}^*) + \epsilon^* \qquad (2.1.1)$$

where $f^*(\boldsymbol{x}, \boldsymbol{\beta}^*)$ consists of all the terms in the polynomial model. For a thorough discussion of RSM, see Myers(1971), Box and Draper (1987), Khuri and Cornell (1996) and Myers and Montgomery (1995). A comprehensive review of this subject is also given by Myers, Khuri, and Carter (1989).

2.2 Multiresponse Experiments

In the early development of response surface methodology, only a single response variable was considered. In many experimental situations, however, the experimenter

is generally interested in more than one response. For example, a chemical engineer may be interested in maximizing the yield (response one) while minimizing the cost (response two) of a certain chemical process. Also, an agronomist may want to maximize the harvest yield and minimize fertilizer cost for a certain type of crop. Experiments in which more than one response is measured and studied for a given setting of the experimental factors (input variables) are called *multiresponse experiments*.

The analysis of multiresponse experiments is quite different from that of single-response experiments. When analyzing data from a multiresponse experiment, one must take into account possible interrelationships that exist among the responses of interest. Thus the multivariate nature of the data obtained from multiresponse experiments must be recognized and considered in the analysis.

An important objective in any multiresponse experiment is the estimation of the relationships between the responses and the input variables within a given region of interest. Suppose that there are r responses $y_1, y_2, \ldots, y_r$, and k input variables $x_1, x_2, \ldots, x_k$ of interest. Let $\{y_{iu}\}$, $i = 1, 2, \ldots, r$, $u = 1, 2, \ldots, n$, represent n sets of observations on each of the r response variables. Then the relationship between the responses and the input variables, as given in (2.1.1), can be represented by

$$y_{iu} = f_i^*(\boldsymbol{x}_u, \boldsymbol{\beta}_i) + \epsilon_{iu}, \quad i = 1, 2, \ldots, r, \quad u = 1, 2, \ldots, n, \tag{2.2.1}$$

where $\boldsymbol{x}_u = (x_{u1}, x_{u2}, \ldots, x_{uk})'$ is the vector of input variables at the uth setting, $\boldsymbol{\beta}_i$ is the vector of unknown parameters associated with the ith response, ϵ_{iu} is a random error, and f_i^* is a response function which can be represented by a polynomial of degree d.

The estimation of the relationships in (2.2.1) involves finding estimates for the parameters $\boldsymbol{\beta}_i$, $i = 1, 2, \ldots, r$. However, these estimates require the collection of observations for the responses based on given values of the input variables. The values of $\boldsymbol{x}_u$ wherein the observed response values will be attained, that is, the design points, are determined according to an appropriate criterion. Once a response surface design

has been properly selected according to some criterion, data are collected and are used to obtain estimates of the unknown parameters. The generalized least squares method or the maximum likelihood method can be used to obtain these estimates to produce the fitted model. If the fitted models are deemed adequate, then they are used to determine the conditions or settings of the input variables that result in estimated optimal values of the mean responses.

2.3 Multiresponse Parameter Estimation

There are two approaches that have been mainly used in the estimation of parameters of multiresponse models. The first one is the generalized least squares method introduced by Zellner (1962). This approach uses his seemingly unrelated regression models. It also involves the estimation of the variance-covariance matrix Σ associated with the r responses. Zellner (1962) showed that the parameter estimates obtained through this method are at least asymptotically more efficient than those obtained using the single-equation application of least squares. A discussion of this method is given in Chapter 3. The other approach developed by Box and Draper (1965), uses a Bayesian argument wherein non-informative prior distributions are assigned to all the parameters in the model, including the variance-covariance matrix Σ. This method is briefly explained in Appendix A. Note that the Box and Draper (1965) approach is more general than Zellner's (1962) method since it applies to nonlinear as well as linear multiresponse models. However, Box et al. (1973) showed that the former method can lead to misleading results if there are linear relationships among the responses. Therefore it is important that these dependencies be removed before the Box and Draper (1965) estimation procedure is used. Box et al. (1973) proposed a procedure for detecting the existence of these dependencies when the multiresponse data are subject to round off errors. An extension of this procedure to situations in

which the response variables have different scales of measurement was introduced by Khuri (1990a)

More recent results on the estimation of parameters in a multiresponse model are given by Kang and Bates (1990) They presented an approximate inference procedure which can be used to obtain approximate confidence regions for the parameters in the multiresponse model Their procedure mainly uses the estimates derived from the Box and Draper (1965) approach

2.4 Multiresponse Optimization

The literature addressing the problem of multiresponse optimization is not as extensive as that for single-response optimization Khuri (1996a) defines multiresponse optimization as the determination of operating conditions on a set of input variables that yield optimal (or near optimal) values for several response variables considered simultaneously This particular problem of simultaneous optimization of responses is addressed in the works done by Harrington (1965), Derringer and Suich (1980), and Khuri and Conlon (1981) See also, Khuri and Cornell (1996 Chapter 7)

Harrington (1965), and later Derringer and Suich (1980), introduced the concept of desirability whereby each response function is transformed to a desirability function This method makes use of a desirability function in which the experimenter's own priorities and desires on the particular responses considered are built into the optimization procedure Harrington (1965) transformed each predicted response function, y_i, to a desirability function d_i^*, where $0 \leq d_i^* \leq 1$, $i = 1, 2, \ldots, r$, such that d_i^* increases as the desirability of the corresponding property which y_i measures increases He proposed the exponential type transformation given by $d_i^* = e^{-|y_i|^s}$ where s is a positive number specified by the user Derringer and Suich (1980) introduced more general transformations which offer the user greater flexibility in setting up the desirability functions Myers and Montgomery (1995, pp 250-251) gave the following

description of this approach One and two-sided desirability functions are used depending on whether the response is to be minimized or maximized or has an assigned target value For the ith response with a target value level A B and C are assigned so that $A \leq B \leq C$ A product is considered unacceptable if $\hat{y}_i < A$ or $\hat{y}_i > C$ The value B is the 'most desirable value' (target value) The quantity d_i^*, the desirability, is defined as

$$d_i^* = \begin{cases} \left(\frac{\hat{y}_i - A}{B - A}\right)^s, & A \leq \hat{y}_i \leq B \\ \left(\frac{\hat{y}_i - C}{B - C}\right)^t, & B \leq \hat{y}_i \leq C \end{cases} \tag{2.4.1}$$

with d_i^* being 0 if $\hat{y}_i > C$ or $\hat{y}_i < A$. The user then decides on the values of s and t according to his/her subjective impression of the total desirability of the response under consideration. In the case where the ith response is to be maximized, the user chooses a value B such that $d_i^* = 1$ for any $\hat{y}_i > B$, that is, the researcher is satisfied if $\hat{y}_i > B$ However the product is assumed to be unacceptable if $\hat{y}_i < A$ In this case $B = C$ and the desirability function is given by

$$d_i^* = \left(\frac{\hat{y}_i - A}{B - A}\right)^s \tag{2.4.2}$$

where $A \leq \hat{y}_i \leq B$ In the case where the ith response is to be minimized, a value of B is chosen such that $\hat{y}_i \leq B$ is considered desirable and produces a $d_i^* = 1$ A value for $\hat{y}_i > C$ is considered unacceptable and thus results in a $d_i^* = 0$ The desirability function is given by

$$d_i^* = \left(\frac{\hat{y}_i - C}{B - C}\right)^t \tag{2.4.3}$$

where $B \leq \hat{y}_i \leq C$

Once the desirability functions for all the r responses are obtained a single composite response is determined This single response is the geometric mean of the d_i^*'s given by $D^* = (\prod_{i=1}^{r} d_i^*)^{1/r}$ The user then needs to find the values of x that maximize D^* The optimization of D^* is fairly simple since D^* is a well-behaved continuous

function A value of D^* close to one implies that all responses are in a desirable range simultaneously Note that the geometric mean is only one possible formulation of D^* Other functions of the d_i^*'s may also be used Examples of these other functions are given by Myers and Montgomery (1995, p 252) For a more detailed discussion and an example of this approach see Myers and Montgomery (1995 pp 247-258)

One advantage of the Derringer and Suich (1980) approach is in the relative ease in which the optimal value of D^* can be obtained However one of its basic characteristics is the subjectivity on the part of the user in defining the desirabilities This is not a problem if the user can correctly evaluate the desirability functions If this is not the case then it is possible that the improperly assessed desirabilities will lead to inaccurate results. Another disadvantage of this approach lies in the fact that it does not take into account the heterogeneity of the variances of the responses or any correlations that might exist among them These are important considerations in a multiresponse optimization procedure

Another multiresponse optimization procedure was introduced by Khuri and Conlon (1981) This procedure is based on the generalized distance approach In this procedure all the response functions of interest are assumed to depend on the same set of of input variables and can be represented by polynomial regression models of the same degree and form within the experimental region Hence this approach applies to linear multiresponse models of the form $\boldsymbol{y}_i = \boldsymbol{X}_{\mathrm{a}}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i$, where $\boldsymbol{X}_{\mathrm{a}}$ is an $n \times p$ matrix of rank p and the values of the ε_i's have means zero, $Var(\varepsilon_i) = \sigma_{ii}\boldsymbol{I}_n$ and $cov(\varepsilon_i, \varepsilon_j) = \sigma_{ij}\boldsymbol{I}_n$, $i \neq j$ $i, j = 1, 2 \quad , r$ Let $\boldsymbol{\Sigma}$ be an $r \times r$ matrix whose (i, j)th element is σ_{ij} $i, j = 1, 2 \quad , r$ Then an estimate of $\boldsymbol{\Sigma}$ is given by

$$\boldsymbol{\Sigma} = \frac{\boldsymbol{Y}'[\boldsymbol{I}_n - \boldsymbol{X}_{\mathrm{a}}(\boldsymbol{X}_{\mathrm{a}}'\boldsymbol{X}_{\mathrm{a}})^{-1}\boldsymbol{X}_{\mathrm{a}}']\boldsymbol{Y}}{n - p} \tag{2.4.4}$$

where $\boldsymbol{Y} = [\boldsymbol{y}_1 \mid \boldsymbol{y}_2 \quad \boldsymbol{y}_r]$ Without loss of generality it can be assumed that the r responses are not linearly related The prediction equation for the ith response is

$\hat{y}_i = \boldsymbol{x}'(\boldsymbol{x})\hat{\boldsymbol{\beta}}_i$, $i = 1, 2, \quad r$ where $\boldsymbol{x}'(\boldsymbol{x})$ is a vector of the same form as a row of the matrix $\boldsymbol{X}_0$ evaluated at the point $\boldsymbol{x}$ $\hat{\boldsymbol{\beta}}_i$ is the OLS estimator of $\boldsymbol{\beta}_i$, and is equal to $(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0'\boldsymbol{y}_i$, $i = 1\ 2 \quad r$ (it can be shown that this estimator coincides with the generalized least-squares estimator of β_i) Let $\hat{\boldsymbol{y}} = (\hat{y}_1, \hat{y}_2 \quad , \hat{y}_r)'$ The variance-covariance matrix of $\hat{\boldsymbol{y}}(\boldsymbol{x})$ at a point $(\boldsymbol{x})$ in the experimental region is given by

$$Var[\hat{\boldsymbol{y}}(\boldsymbol{x})] = \boldsymbol{x}'(\boldsymbol{x})(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{x}'(\boldsymbol{x})\boldsymbol{\Sigma} \tag{2.4.5}$$

An unbiased estimator, $\widehat{Var}[\hat{\boldsymbol{y}}(\boldsymbol{x})]$ of $Var[\hat{\boldsymbol{y}}(\boldsymbol{x})]$ can be obtained by using $\boldsymbol{\Sigma}$ in place of $\boldsymbol{\Sigma}$ in the above equation.

Let ϕ_i be the optimum value of $\hat{y}_i$ optimized individually over the experimental region and let $\boldsymbol{\phi} = (\phi_1, \phi_2, \quad , \phi_r)'$ If these individual optima are attained at the same set $\boldsymbol{x}$, of operating conditions, then an ideal optimum is said to be achieved However such an ideal optimum rarely exists. In more general situations one might want to consider finding compromise conditions on the input variables that are somewhat favorable to all responses This means finding conditions under which the multiresponse function deviates as little as possible from the ideal optimum that is at such conditions a 'near' optimum is achieved for each of the r predicted response functions Such a deviation of the compromising conditions from the ideal optimum can be formulated by means of a distance function which measures the distance between $\hat{\boldsymbol{y}}(\boldsymbol{x})$ and $\boldsymbol{\phi}$ Let this distance function be represented by $\rho[\hat{\boldsymbol{y}}(\boldsymbol{x}), \boldsymbol{\phi}]$ The multiresponse optimization then involves finding conditions on $\boldsymbol{x}$ that minimize this distance function over the experimental region Khuri and Conlon (1981) gave several forms for this distance function One of them is given by the weighted distance

$$\rho[\hat{\boldsymbol{y}}(\boldsymbol{x}), \boldsymbol{\phi}] = [(\hat{\boldsymbol{y}}(\boldsymbol{x}) - \boldsymbol{\phi})'\{Var[\hat{\boldsymbol{y}}(\boldsymbol{x})]\}^{-1}(\hat{\boldsymbol{y}}(\boldsymbol{x}) - \boldsymbol{\phi})]^{1/2} \tag{2.4.6}$$

Since $\boldsymbol{\Sigma}$ is unknown $\widehat{Var}[\hat{\boldsymbol{y}}(\boldsymbol{x})]$ can be used in place of $Var[\hat{\boldsymbol{y}}(\boldsymbol{x})]$ in the above equation. If $\boldsymbol{x}_0$ is the point in the experimental region at which $\rho[\hat{\boldsymbol{y}}(\boldsymbol{x}), \boldsymbol{\phi}]$ attains its

absolute minimum, and if m_0 is the value of this minimum, then at x_0 the experimental conditions can be described as being near optimal for each of the r response functions. The smaller the value of m_0, the closer these conditions are to representing the 'ideal' optimum. A more detailed discussion and an example of this approach is given in Khuri and Cornell (1996, pp 285-293).

An advantage of the Khuri and Conlon (1981) generalized distance procedure over the Derringer and Suich (1980) desirability approach is that the generalized distance approach takes into consideration the correlations among the responses as well as the magnitudes of their variances. A disadvantage of this procedure, however, is the requirement that all the response functions must be represented by polynomial regression models of the same degree and form. In some multiresponse situations this may not be practical. This problem will be addressed in Chapter 3.

Other optimization procedures that involve several responses, but are not concerned with 'simultaneous' optimization, are given by Myers and Carter (1973) and Biles (1975). Myers and Carter (1973) introduced the dual response approach for the optimization of one response with constraints placed on another response. In the dual response approach, the two responses are categorized as 'primary', represented by Y_p and 'secondary', represented by Y_s. The goal of the procedure is to determine conditions on $\boldsymbol{x}$ that optimize the primary response function subject to an equality constraint on the secondary response function. The optimization problem is to

$$optimize \quad Y_p \quad subject \quad to \quad Y_s = c. \tag{2.4.7}$$

In this procedure it is assumed that both responses can be estimated by using second-order models of the form

$$Y_p = b_o + \boldsymbol{x}'\boldsymbol{b} + \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x} + \varepsilon_p \tag{2.4.8}$$

$$Y_s = c_o + \boldsymbol{x}'\boldsymbol{c} + \boldsymbol{x}'\boldsymbol{C}\boldsymbol{x} + \varepsilon_s \tag{2.4.9}$$

where $\boldsymbol{B}$ and $\boldsymbol{C}$ are $k \times k$ matrices, $\boldsymbol{b}$ and $\boldsymbol{c}$ are $k \times 1$ vectors, and b_0 and c_0 are scalars. To restrict the search area to a spherical region of radius ρ^2, an additional constraint given by

$$\boldsymbol{x}'\boldsymbol{x} = \rho^2 \tag{2.4.10}$$

is considered.

To solve this optimization problem, Myers and Carter (1973) showed that a stationary point $\boldsymbol{x}$ has to satisfy

$$(\boldsymbol{B} - \lambda_1 \boldsymbol{C} - \lambda_2 \boldsymbol{I})\boldsymbol{x} = \frac{\lambda_1 \boldsymbol{c} - \boldsymbol{b}}{2} \tag{2.4.11}$$

where λ_1 is the Lagrange multiplier of the constraint in (2.4.7), λ_2 is the Lagrange multiplier of the constraint in (2.4.10). The dual response approach consists of choosing λ_2 so that the matrix on the left hand side of the above equation gives either a maximum or a minimum. This is accomplished by finding the eigenvalues of $\boldsymbol{B} - \lambda_1 \boldsymbol{C}$. The procedure calls for fixing the value of λ_1, computing a suitable value of λ_2, and obtaining $\boldsymbol{x}$ from (2.4.11). By trying different values of λ_1, the experimenter obtains a value of $\boldsymbol{x}$ that simultaneously satisfies (2.4.10) and ensures a satisfactory secondary response. Biles (1975) extended these results to the case where there are several secondary response functions.

Nonlinear optimization procedures can be used to solve the dual response problem. This is due to the fact that usually both the primary and the secondary response functions are quadratic functions. There are several nonlinear optimization procedures available in the operations research (OR) literature. Del Castillo and Montgomery (1993) showed that one particular nonlinear programming algorithm works best in solving the dual response problem. This particular algorithm is the *generalized reduced gradient* (GRG) algorithm. The following is a description of the GRG method (see Del Castillo and Montgomery (1993) p 201). The GRG method optimizes a nonlinear objective function subject to nonlinear equality constraints. Inequality

constraints can be transformed to equality constraints by introducing surplus variables For example, if a constraint has the form $x_1 + x_2 \leq c$ the surplus variable x_s is introduced and the constraint is rewritten as $x_1 + x_2 + x_s = c$ The surplus variable is assigned a coefficient of zero in the objective function which is the function to be optimized

Given a current feasible region, the GRG method tries to find a direction of movement that will improve the value of the objective function Then it tries to find an appropriate step size for moving in this direction while checking to see that the constraints are not violated The direction of the movement is given by the *reduced gradient*. The negative of the reduced gradient gives the direction of steepest ascent for a maximization problem and steepest descent for a minimization problem It is called a reduced gradient because of the way the variables are grouped into two sets by the algorithm at each iteration One set consists of basic (dependent) variables These are the variables whose current values are *strictly between* their bounds The other set, called the nonbasic (independent) variables, consists of the remaining variables whose current values are at their bounds

If there are m constraints and k variables, usually $m < k$, the system of equations generated by the constraints will have more unknowns than equations and cannot be solved explicitly However the system can be solved for m of the basic variables in terms of the remaining $k - m$ nonbasic variables If the constraint functions are 'well behaved', solving this system can be done by numerical methods The solution of the system will be feasible and will have at most $k - m$ nonbasic variables The gradient of the objective function with respect to the nonbasic variables is called the *reduced gradient*. If the magnitude of the reduced gradient at the current point is small, then the algorithm stops Otherwise, a line search in the direction of the negative of the reduced gradient is performed a new point is found and the process is repeated

The advantages of the GRG method as discussed by Del Castillo and Montgomery (1993) are as follows It allows

1 more constraints, hence response functions, to be considered

2 the analyst to directly specify the radius of the spherical region of interest in the dual response approach rather than indirectly specifying this through the Lagrange multipher λ_l

3 for other forms of the constraints that define the region of interest

From these advantages one can conclude that the GRG method is more flexible than the Myers and Carter (1973) procedure and that it allows for the handling of more than two responses For examples of the GRG approach see Del Castillo and Montgomery (1993 pp 202-204)

2.5 Blocking in RSM

Blocking is an experimental design technique which can be used in experiments that are too large to allow all runs to be made under homogeneous conditions In RSM, there are many situations where this is the case For example a block may refer to a group of experiments performed on the same day but several days of experiments are desired, or a group of experiments for which it was possible to use the same batch of raw material but several batches of raw material are needed Because of the possibility of running large experiments under non-homogeneous conditions, it is important to consider designs that can be run in blocks Box and Hunter (1957) indicated however that the manner in which the experimental points of a particular design are assigned to the blocks greatly affects the efficiency of the estimation of model parameters and the ease with which these estimates are calculated To solve this problem they introduced the concept of *orthogonal blocking* for second order designs. If a design blocks orthogonally, then the inclusion of block effects in the

model does not affect the estimates of the model parameters associated with the model's polynomial effects. The intercept estimate, however, depends on the blocks. The analysis of the experiment then proceeds exactly as if there were no blocking except for a reduction in the residual sum of squares due to the removal of the blocks sums of squares. Box and Hunter (1957, 1961a, b) developed the general conditions that give rise to orthogonal blocking in the second order case. They also presented orthogonal blocking arrangements for some central composite designs (CCD). Box and Hunter (1957) noted that one aspect which makes these orthogonal blocking arrangements particularly useful arises out of the nature in which CCDs are used, that is, in the exploration of response surfaces. trials are usually performed sequentially and often have as their objective the optimization of some response. For example, suppose that four variables were studied with the intention of finding an optimum set of conditions for a new process using the method of steepest ascent. The method of steepest ascent is an optimization technique which involves sequential movement from one region to another, to search for conditions in which the process output value is improved. In this case, a half replicate of a 2^4 factorial design with two center point replicates (block 1) could first be run. This design would supply estimates of the four first-order terms, combinations of the interaction terms and of the sum of the four quadratic terms. If the values of the coefficient estimates of the first-order terms were very large compared to their standard errors, progress could probably be made without a more elaborate design by moving in the indicated direction of increasing yield. Continuous movement would be made until improvement in that direction was exhausted. At this point a new first-order design would be started at the improved set of conditions. If the first order terms were small, or a tentative move had not met with success, the second half of the factorial design together with two additional center point replicates (block 2) could be carried out, and the situation is reviewed in the light of the set of 20 trials which are now available. Finally, if

evidence indicates that further progress can be attained only by fitting the second-degree equation, the third block of ten points consisting of the eight axial points and two center points would be added. The complete set of thirty trials could then be used to provide efficient estimates of the best fitting second-order model and further progress would follow. Thus, historically, blocking in RSM is mainly used as a tool in sequential experimentation. For a more detailed discussion of orthogonal blocking for second-order designs, see Box and Draper (1987, pp 509-520), Khuri and Cornell (1996, Chapter 8) and Myers and Montgomery (1995, pp 328-341). Results on non-orthogonal blocking for second order models are also given by Dey and Das (1970) and Khuri and Cornell (1996, Chapter 8).

Some of the more current results on blocking in RSM are given by Cook and Nachtsheim (1989), Adhikary and Panda (1990), Khuri (1992, 1994, 1996b) and Khuri and Cornell (1996, Chapter 8). Cook and Nachtsheim (1989) introduced computer-aided blocking of factorial and response-surface designs while Adhikary and Panda (1960) presented a sequential method for constructing second order rotatable designs in non-orthogonal blocks. Khuri(1992) considered response-surface models with random block effects and generalized the result given by Box and Hunter (1957) on the general conditions for response surface designs to block orthogonally. Details of his work are given in Chapter 5. Khuri (1994) also presented results on the effect of blocking on the estimation of the mean response, on the prediction variance and on the optimum of the response surface. These results are described in Chapter 4. More recently, Khuri (1996b) described results on response surface models with mixed effects. A description of these are also given in Chapter 5.

2.6 Multiresponse Design Criteria

In selecting a design, most researchers rely on a design optimality criterion. Most design optimality criteria for single-response experiments are based on some of the

properties of the $\boldsymbol{X}'\boldsymbol{X}$ matrix for a design matrix $\boldsymbol{X}$. One such criterion is the D-optimality criterion. D-optimality is one of the most popular design criterion. D optimal designs maximize the determinant of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$. Its appeal is based on the property that the determinant of $\boldsymbol{X}'\boldsymbol{X}$ is inversely proportional to the square of the volume of the confidence region (hyperellipsoid) on the model parameters. Thus a large determinant of $\boldsymbol{X}'\boldsymbol{X}$ or equivalently a small determinant of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ implies better estimation of the model parameters.

In the multiresponse case, one needs to consider criteria that involve all the response variables. One such criterion was developed by Fedorov (1972). Fedorov (1972) introduced a sequential procedure that can be used to obtain multiresponse D-optimal designs. However, his procedure requires knowing $\boldsymbol{\Sigma}$, the variance-covariance matrix associated with the r responses. Wijesinha and Khuri (1987a) modified Fedorov's procedure by replacing $\boldsymbol{\Sigma}$ with Zellner's (1962) estimate. Another design criterion presented by Wijesinha and Khuri (1987b) is one based on maximizing the power of the multivariate lack of fit test introduced by Khuri (1985). This criterion is an extension of the method used by Jones and Mitchell (1978) in the single-response case. Khuri (1990b) introduced the concept of multiresponse rotatability, based on the work done by Box and Hunter (1957). The concept of rotatability was first introduced by Box and Hunter (1957) and has since become one of the more popular and important design criteria. With a rotatable design, the prediction variance at a point $\boldsymbol{x}$ is a function only of the distance between the point $\boldsymbol{x}$ and the center of the design. This implies that with a rotatable design, the prediction variance is the same at all points $\boldsymbol{x}$ that are equidistant from the design center. Khuri (1990b) extended this property to the multiresponse case. He defined multiresponse rotatability as a design property under which the variances and covariances among the predicted responses are constant at all points that are equidistant from the design center. He showed that multiresponse rotatability can be achieved if and only if the design is rotatable for

a single-response model that has the highest degree among all the response models under consideration

In choosing a response surface design, an experimenter may want to consider a design criterion that takes into account both the bias and the variance of the fitted model The integrated mean-squared error (IMSE) criterion, introduced by Box and Draper (1959), considers both properties The IMSE criterion is also known as the J-criterion

The J- criterion is described as follows (see Khuri and Cornell (1996), Chapter 6) Suppose that the fitted model for the mean response is given by $y = \boldsymbol{X}_1^*\boldsymbol{\beta}_1 + \epsilon$ of order d_1 Let the true model for the mean response be $\eta = \boldsymbol{X}_1^*\boldsymbol{\beta}_1 + \boldsymbol{X}_2^*\boldsymbol{\beta}_2$ of order d_2 ($d_2 > d_1$) Then, at each point $\boldsymbol{x}$ in the experimental region $\Re$ the mean squared error (MSE) of $y(\boldsymbol{x})$ is $MSE[y(\boldsymbol{x})] = E[y(\boldsymbol{x}) - \eta(\boldsymbol{x})]^2$ where $y(\boldsymbol{x})$ is the least squares estimate of the mean response value, $\eta(\boldsymbol{x})$ at $\boldsymbol{x}$ obtained with the fitted model If this quantity is averaged over the region $\Re$ we obtain. $\Omega \int_{\Re} E[\hat{y}(\boldsymbol{x}) - \eta(\boldsymbol{x})]^2 d\boldsymbol{x}$ where $\Omega^{-1} = \int_{\Re} d\boldsymbol{x}$ Now in order to be able to center our attention on designs which may or may not contain a fixed number of points and also to remove the dependence of the mean squared error of $y(\boldsymbol{x})$ on the value of the variance σ^2, the criterion is put on a 'per observation' basis and is expressed as $J = (N\Omega/\sigma^2) \int_{\Re} E[y(\boldsymbol{x}) - \eta(\boldsymbol{x})]^2 d\boldsymbol{x}$ The difference $y(\boldsymbol{x}) - \eta(\boldsymbol{x})$ can be partitioned as $\bar{y}(\boldsymbol{x}) - \eta(\boldsymbol{x}) = \{\bar{y}(\boldsymbol{x}) - E[\bar{y}(\boldsymbol{x})]\} + \{E[\bar{y}(\boldsymbol{x})] - \eta(\boldsymbol{x})\}$ This partitioning enables us to separate J into two parts namely,

$$\begin{aligned} J &= (N\Omega/\sigma^2) \int_{\Re} Var[y(\boldsymbol{x})] d\boldsymbol{x} + (N\Omega/\sigma^2) \int_{\Re} \{E[y(\boldsymbol{x})] - \eta(\boldsymbol{x})\}^2 d\boldsymbol{x} \\ &= V + B \end{aligned}$$

The quantities V and B are called the average variance of $y(\boldsymbol{x})$ and the average squared bias of $y(\boldsymbol{x})$ respectively The J-criterion is concerned with minimizing V and B but regards the minimization of B to be more important Kiefer (1972)

Introduced a multiresponse version of the integrated mean-squared error (IMSE) criterion. He presented an iterative procedure for minimizing the IMSE based on the results given by Karson, Manson and Hader (1969) under the assumption that the variance-covariance matrix Σ associated with the r responses is known. More recently, Kim and Draper (1994) presented results for minimizing the IMSE for two responses assuming also that the variance-covariance matrix Σ is known. A more detailed discussion of this criterion is given in Chapter 6. Examples and applications of the J criterion can be found in Khuri and Cornell (1996, pp.207-231) and Myers and Montgomery (1995, pp 406-421).

CHAPTER 3

OPTIMIZATION OF MULTIRESPONSE FUNCTIONS

3.1 Introduction

One of the objectives of response surface methodology is the determination of conditions on the input variables that produce optimum values of the predicted response. In this chapter we present a method that can be used to find these optimum conditions in the multiresponse case. As mentioned in Chapter 1, the importance of multiresponse optimization can be seen in many practical situations. For example in the pharmaceutical area, the scientist is not only concerned with maximization of the efficacy of a drug but also with the minimization of undesirable side effects of the drug. The optimization of a multiresponse function must take into account any interrelationships that exist among the responses of interest. Failure to do so may lead to optimal conditions for one response that are not optimal for the other responses.

The simultaneous optimization of several response functions was addressed by Khuri and Conlon (1981) using what is called the generalized distance approach. A description of this approach was given in Chapter 2. In this optimization procedure, Khuri and Conlon (1981) considered the case in which the responses are represented by polynomial regression models of the same degree and form. It is not always the case that models of the same degree and form are appropriate for all the responses. In some instances, the design chosen to run the multiresponse experiment may not

provide a sufficient number of runs or points to support fitting a higher degree model that some of the responses (but not all) demand It is also possible that the best model for one response may not be the best' for the other responses. Thus we extend the generalized distance approach presented by Khuri and Conlon (1981) to the case where the polynomial regression models are not necessarily of the same degree or form

3.2 The Multiresponse Model

In this section we will establish a single model that is a multiresponse model which will represent all the responses of interest The true form of each response function is assumed to be adequately approximated by a polynomial model Once the multiresponse model is established estimates of the model parameters are derived by using the generalized least-squares method

Suppose that there are r responses of interest namely $y_1, y_2, \ldots, y_r$, each of which depends on the same set of k independent variables denoted by $x_1, x_2, \ldots, x_k$ within an experimental region $\mathcal{R}$. Using the results given by Khuri (1990b), the model for the ith response is

$$y_i(\boldsymbol{x}) = \mu_i(\boldsymbol{x}) + \epsilon_i \quad i = 1, 2, \ldots, r \tag{3.2.1}$$

where $\mu_i(\boldsymbol{x})$ is a polynomial of degree d_i defined at $\boldsymbol{x} = (x_1, x_2, \ldots, x_k)'$ and ϵ_i is a random error. Model (3.2.1) can be expressed as

$$y_i(\boldsymbol{x}) = \boldsymbol{f}_i'(\boldsymbol{x})\boldsymbol{\gamma}_i + \epsilon_i, \quad i = 1, 2, \ldots, r \tag{3.2.2}$$

where $\boldsymbol{f}_i'(\boldsymbol{x})$ is a vector of q_i elements consisting of powers and cross-products of powers of $x_1, x_2, \ldots, x_k$ up to degree d_i and whose first element is equal to one, and $\boldsymbol{\gamma}_i$ is a vector of unknown parameters of order $q_i \times 1$

If observations are made on the r responses at n design settings then from (3.2.2) we obtain the models

$$\boldsymbol{y}_i = \boldsymbol{U}_i\boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i, \quad i = 1, 2, \ldots, r \tag{3.2.3}$$

where $\boldsymbol{y}_i$ and $\boldsymbol{\epsilon}_i$ are the vectors of observations and random errors respectively, for the ith response, and $\boldsymbol{U}_i$ is of order $n \times q_i$ and rank q_i whose first column is $\mathbf{1}_n$. The models in (3.2.3) can be expressed as a single multiresponse model of the form

$$\boldsymbol{y} = \boldsymbol{U}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \tag{3.2.4}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1', \boldsymbol{y}_2', \ldots, \boldsymbol{y}_r')'$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2', \ldots, \boldsymbol{\gamma}_r')'$, and $\boldsymbol{U} = diag(\boldsymbol{U}_1, \boldsymbol{U}_2, \ldots, \boldsymbol{U}_r)$. The variance-covariance matrix of $\boldsymbol{\epsilon}$ is $\boldsymbol{\Sigma} \otimes \boldsymbol{I}_n$, where $\boldsymbol{\Sigma} = (\sigma_{ij})$ is the variance-covariance matrix of the r responses. The best linear unbiased estimator of $\boldsymbol{\gamma}$ is

$$\boldsymbol{\gamma} = [\boldsymbol{U}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{U}]^{-1}\boldsymbol{U}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} \tag{3.2.5}$$

The variance-covariance matrix of $\boldsymbol{\gamma}$ is

$$Var(\boldsymbol{\gamma}) = [\boldsymbol{U}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{U}]^{-1} \tag{3.2.6}$$

Let $\boldsymbol{y}(\boldsymbol{x})$ be the $r \times 1$ vector of predicted responses at a point $\boldsymbol{x}$ in the experimental region, that is,

$$\boldsymbol{y}(\boldsymbol{x}) = [y_1(\boldsymbol{x}), y_2(\boldsymbol{x}), \ldots, y_r(\boldsymbol{x})]' \tag{3.2.7}$$

where $y_i(\boldsymbol{x}) = \boldsymbol{f}_i'(\boldsymbol{x})\boldsymbol{\gamma}_i$ and $\boldsymbol{\gamma}_i$ is the best linear unbiased estimator of $\boldsymbol{\gamma}_i$ obtained from (3.2.5). Formula (3.2.7) can then be written as

$$\hat{\boldsymbol{y}}(\boldsymbol{x}) = \boldsymbol{\Lambda}'(\boldsymbol{x})\boldsymbol{\gamma}, \tag{3.2.8}$$

where $\boldsymbol{\Lambda}'(\boldsymbol{x}) = diag(\boldsymbol{f}_1'(\boldsymbol{x}), \boldsymbol{f}_2'(\boldsymbol{x}), \ldots, \boldsymbol{f}_r'(\boldsymbol{x}))$, and $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2', \ldots, \boldsymbol{\gamma}_r')'$. From (3.2.6) and (3.2.8) the variance-covariance matrix of $\hat{\boldsymbol{y}}(\boldsymbol{x})$ is given by

$$Var[\hat{\boldsymbol{y}}(\boldsymbol{x})] = \boldsymbol{\Lambda}'(\boldsymbol{x})[\boldsymbol{U}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{U}]^{-1}\boldsymbol{\Lambda}(\boldsymbol{x}) \tag{3.2.9}$$

The matrix $\boldsymbol{\Sigma}$ is usually unknown To estimate the elements of $\boldsymbol{\Sigma}$ we use the consistent estimator given by Zellner (1962) [see also Srivastava and Giles (1987 pp 13-17)] This estimator is $\boldsymbol{S} = (s_{ij})$ where

$$s_{ij} = (1/n)\boldsymbol{y}_i'(\boldsymbol{I} - \boldsymbol{U}_i(\boldsymbol{U}_i'\boldsymbol{U}_i)^{-1}\boldsymbol{U}_i')(\boldsymbol{I} - \boldsymbol{U}_j(\boldsymbol{U}_j'\boldsymbol{U}_j)^{-1}\boldsymbol{U}_j')\boldsymbol{y}_j \tag{3 2 10}$$

Note that other estimators of $\boldsymbol{\Sigma}$ may be used A brief description of the maximum likelihood estimator of $\boldsymbol{\Sigma}$ is given in Appendix A However this approach was developed under the Bayesian framework and also under the assumption that $\boldsymbol{\epsilon}$ follows a multivariate normal distribution The assumption of normality for $\boldsymbol{\epsilon}$ in our case is extraneous in the sense that we do not need it for the optimization procedure that will be presented in the next section Hence, for our optimization procedure we will use $\boldsymbol{S}$ to estimate $\boldsymbol{\Sigma}$

Using $\boldsymbol{S}$ given in (3 2 10) $\boldsymbol{\gamma}$ is estimated by

$$\hat{\boldsymbol{\gamma}}_e = [\boldsymbol{U}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{U}]^{-1}\boldsymbol{U}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} \tag{3 2 11}$$

and the corresponding vector of predicted responses is given by

$$\boldsymbol{y}_e(\boldsymbol{x}) = \boldsymbol{\Lambda}'(\boldsymbol{x})\hat{\boldsymbol{\gamma}}_e \tag{3 2 12}$$

Thus, we approximately have

$$Var(\boldsymbol{y}_e(\boldsymbol{x})) = \boldsymbol{\Lambda}'(\boldsymbol{x})[\boldsymbol{U}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{U}]^{-1}\boldsymbol{\Lambda}(\boldsymbol{x}) \tag{3 2 13}$$

The assumed nonsingularity of $\boldsymbol{\Sigma}$ does not ensure the nonsingularity of $\boldsymbol{S}$ Srivastava and Giles (1987 p 16) showed that if the number of responses, r is greater than the number of observations n then $\boldsymbol{S}$ will be singular with probability one They also showed that $r \leq n$ is a necessary but not sufficient condition for the nonsingularity of $\boldsymbol{S}$

3.3 Procedures for the Simultaneous Optimization of the Responses

In the previous section we have obtained the estimators of the model parameters. This allows us to proceed with the simultaneous optimization of the responses via the generalized distance approach.

Multiresponse optimization under the generalized distance approach amounts to finding conditions on $\boldsymbol{x}$ that minimize the distance between the predicted responses and their corresponding optimum values. Let ϕ_i be the optimum value of $\hat{y}_{ie}(\boldsymbol{x})$ optimized individually over an experimental region $\mathcal{R}$, where $\hat{y}_{ie}(\boldsymbol{x})$ is the ith element of $\hat{\boldsymbol{y}}_e(\boldsymbol{x})$ in (3.2.12). Let $\boldsymbol{\phi} = (\phi_1 \ \phi_2 \ \ldots \ \phi_r)'$. Also, let $\rho[\hat{\boldsymbol{y}}_e(\boldsymbol{x}), \boldsymbol{\phi}]$ be a distance function which measures the distance of $\hat{\boldsymbol{y}}_e(\boldsymbol{x})$ from $\boldsymbol{\phi}$. Consider the distance function given by

$$\rho[\hat{\boldsymbol{y}}_e(\boldsymbol{x}), \boldsymbol{\phi}] = [(\hat{\boldsymbol{y}}_e(\boldsymbol{x}) - \boldsymbol{\phi})'(Var(\hat{\boldsymbol{y}}_e(\boldsymbol{x})))^{-1}(\hat{\boldsymbol{y}}_e(\boldsymbol{x}) - \boldsymbol{\phi})]^{1/2}$$

which can also be written as

$$\rho[\hat{\boldsymbol{y}}_e(\boldsymbol{x}), \boldsymbol{\phi}] = \{(\hat{\boldsymbol{y}}_e(\boldsymbol{x}) - \boldsymbol{\phi})'[\mathbf{A}(\boldsymbol{x})(\boldsymbol{U}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{U})^{-1}\mathbf{A}(\boldsymbol{x})]^{-1}(\hat{\boldsymbol{y}}_e(\boldsymbol{x}) - \boldsymbol{\phi})\}^{1/2} \quad (3.3.1)$$

To minimize $\rho[\hat{\boldsymbol{y}}_e(\boldsymbol{x}), \boldsymbol{\phi}]$, we need to take the variability of $\boldsymbol{\phi}$ which is a random vector into account. To do so, we apply the following procedure given by Khuri and Conlon (1981). Let ζ_i be the optimum value of the true mean of the ith response optimized individually over $\mathcal{R}$ and let $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \ldots \ \zeta_r)'$. The objective is to find an $\boldsymbol{x}$ in $\mathcal{R}$ such that $\rho[\hat{\boldsymbol{y}}_e(\boldsymbol{x}), \boldsymbol{\zeta}]$ has an absolute minimum over $\mathcal{R}$. However, ρ is a function of $\boldsymbol{\zeta}$ which is unknown. As a result, it is impossible to minimize ρ directly. In this case, we will minimize an upper bound instead. This upper bound can be established by setting up a confidence region on $\boldsymbol{\zeta}$. Let $\mathcal{D}_{\boldsymbol{\zeta}}$ be a confidence region for $\boldsymbol{\zeta}$, then whenever $\boldsymbol{\zeta} \in \mathcal{D}_{\boldsymbol{\zeta}}$,

$$\rho[\hat{\boldsymbol{y}}_e(\boldsymbol{x}), \boldsymbol{\zeta}] \leq max_{\boldsymbol{\varphi} \in \mathcal{D}_{\boldsymbol{\zeta}}} \rho[\hat{\boldsymbol{y}}_e(\boldsymbol{x}), \boldsymbol{\varphi}] \quad (3.3.2)$$

Since $\boldsymbol{\zeta}$ is unknown, we adopt a conservative approach and minimize the function on the right hand side of the above inequality over $\mathcal{R}$. The following result is needed for this approach.

Theorem 3.3.1. The confidence regions given by

$$g_{1i} \le \zeta_i \le g_{2i}, \quad i = 1, 2, \ldots, r \tag{3.3.3}$$

hold simultaneously to produce a rectangular confidence region with an approximate confidence coefficient of at least $1 - \alpha' = 1 - r\alpha$, *where*

$$g_{1i} = \phi_i - \mathcal{L}_i(\boldsymbol{U}, \xi_i, \boldsymbol{S})z_{\alpha/2}, \quad g_{2i} = \phi_i + \mathcal{L}_i(\boldsymbol{U}, \tilde{\xi}_i, \boldsymbol{S})z_{\alpha/2} \tag{3.3.4}$$

In (3.3.4), $\mathcal{L}_i(\boldsymbol{U}, \xi_i, \boldsymbol{S}) = [\boldsymbol{f}_i'(\boldsymbol{\xi}_i)\boldsymbol{V}^{ii}\boldsymbol{f}_i(\boldsymbol{\xi}_i)]^{1/2}$ *where* $\boldsymbol{\xi}_i$ *is the point at which* $y_{i\ell}(\boldsymbol{x})$ *attains its individual optimum* ϕ_i, $\boldsymbol{f}_i'(\boldsymbol{\xi}_i)$ *is the vector* $\boldsymbol{f}_i'$ *evaluated at* $\boldsymbol{\xi}_i$, $\boldsymbol{V}^{ii}$ *is the* i*th diagonal submatrix of*

$$\begin{pmatrix} s^{11}\boldsymbol{U}_1'\boldsymbol{U}_1 & s^{12}\boldsymbol{U}_1'\boldsymbol{U}_2 & \cdots & s^{1r}\boldsymbol{U}_1'\boldsymbol{U}_r \\ s^{21}\boldsymbol{U}_2'\boldsymbol{U}_1 & s^{22}\boldsymbol{U}_2'\boldsymbol{U}_2 & \cdots & s^{2r}\boldsymbol{U}_2'\boldsymbol{U}_r \\ \vdots & \vdots & \ddots & \vdots \\ s^{r1}\boldsymbol{U}_r'\boldsymbol{U}_1 & s^{r2}\boldsymbol{U}_r'\boldsymbol{U}_2 & \cdots & s^{rr}\boldsymbol{U}_r'\boldsymbol{U}_r \end{pmatrix}^{-1} = [\boldsymbol{U}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{U}]^{-1}$$

and s^{ij} *is the* ij*th element of* $\boldsymbol{S}^{-1}$. *This rectangular confidence region can be used in place of* $\mathcal{D}_{\boldsymbol{\zeta}}$ *in (3.3.2).*

Proof. Let $\boldsymbol{\xi}_i$ denote the value of $\boldsymbol{x}$ at which the true mean of the ith response achieves an optimum ζ_i over $\mathcal{R}$, ignoring the remaining responses. An unbiased estimator of ζ_i is

$$\zeta_i = \boldsymbol{f}_i'(\boldsymbol{\xi}_i)\boldsymbol{\gamma}_i, \quad i = 1, 2, \ldots, r,$$

where $\boldsymbol{f}_i'(\boldsymbol{\xi}_i)$ is the vector $\boldsymbol{f}_i'$ evaluated at $\boldsymbol{\xi}_i$, for which the variance is

$$Var(\zeta_i) = \boldsymbol{f}_i'(\boldsymbol{\xi}_i)\boldsymbol{V}^{ii}\boldsymbol{f}_i(\boldsymbol{\xi}_i), \quad i = 1, 2, \ldots, r$$

where $\boldsymbol{V}^{ii}$ is the ith diagonal submatrix of $[\boldsymbol{U}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{U}]^{-1}$ written in the form

$$[\boldsymbol{U}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{U}]^{-1} = \begin{pmatrix} \sigma^{11}\boldsymbol{U}_1'\boldsymbol{U}_1 & \sigma^{12}\boldsymbol{U}_1'\boldsymbol{U}_2 & & \sigma^{1r}\boldsymbol{U}_1'\boldsymbol{U}_r \\ \sigma^{21}\boldsymbol{U}_2'\boldsymbol{U}_1 & \sigma^{22}\boldsymbol{U}_2'\boldsymbol{U}_2 & & \sigma^{2r}\boldsymbol{U}_2'\boldsymbol{U}_r \\ & & & \\ \sigma^{r1}\boldsymbol{U}_r'\boldsymbol{U}_1 & \sigma^{r2}\boldsymbol{U}_r'\boldsymbol{U}_2 & & \sigma^{rr}\boldsymbol{U}_r'\boldsymbol{U}_r \end{pmatrix}^{-1}$$

Let

$$T_i = \frac{\zeta_{ie} - \zeta_i}{[\boldsymbol{f}_i'(\boldsymbol{\xi}_i)\boldsymbol{V}^{ii}\boldsymbol{f}_i(\boldsymbol{\xi}_i)]^{\frac{1}{2}}}$$

where $\boldsymbol{V}^{ii}$ is the ith diagonal submatrix of $[\boldsymbol{U}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{U}]^{-1}$ $\zeta_{ie} = \boldsymbol{f}_i'(\boldsymbol{\xi}_i)\boldsymbol{\gamma}_{ie}$, and $\boldsymbol{\gamma}_{ie}$ is obtained from equation (3 2 11). Then, using the results of Zellner (1962, p.353), T_i has the standard normal or Z-distribution asymptotically. Now, $P(|T_i| \leq z_{\alpha/2})$ is approximately equal to $1-\alpha$ $i = 1, 2, \ldots, r$. Using the Bonferroni inequality, we have approximately

$$P(\cap_{i=1}^{r}\{|T_i| \leq z_{\alpha/2}\}) \geq 1 - r\alpha$$

If we let $1 - r\alpha = 1 - \alpha^*$, then the inequalities

$$\psi_{1i} \leq \zeta_i \leq \psi_{2i} \quad i = 1, 2, \ldots, r$$

hold simultaneously with an approximate confidence coefficient of at least $1 - \alpha^*$, where

$$\psi_{1i} = \zeta_{ie} - [\boldsymbol{f}_i'(\boldsymbol{\xi}_i)\boldsymbol{V}^{ii}\boldsymbol{f}_i(\boldsymbol{\xi}_i)]^{1/2} z_{\alpha/2},$$

$$\psi_{2i} = \zeta_{ie} + [\boldsymbol{f}_i'(\boldsymbol{\xi}_i)\boldsymbol{V}^{ii}\boldsymbol{f}_i(\boldsymbol{\xi}_i)]^{1/2} z_{\alpha/2}$$

If $\boldsymbol{\xi}_i$ is replaced by $\hat{\boldsymbol{\xi}}_i$, then the above inequalities become inequalities (3 3 3). It follows that an approximate confidence region for $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \ldots, \zeta_r)'$ with confidence coefficient of at least $1 - \alpha^*$ is given by the rectangular region determined by inequalities (3.3 3).

3.4 Numerical Examples

In this section two examples each illustrating the direct minimization of the distance function g, are discussed. The first example uses the results by Tseo et al (1983) (see also Khuri and Cornell 1996 p 295) and deals with four response variables and three input variables. The second example uses the study made by Richert et al (1974) which consists of three response variables and five input variables. For these two examples we show that the 'best' models in terms of the R^2, adjusted R^2, and the Cp criteria are not the same for all the responses. Hence the procedure described in Section 3 3 may be applied to determine the simultaneous optima of the responses.

Example 1. Tseo et al (1983) investigated the effects of washing temperature (x_1), washing time (x_2), and washing ratio (x_3) on springiness (y_1), thiobarbituric acid number (y_2), percent cooking loss (y_3), and whiteness index (y_4) for minced mullet flesh. It is of interest to simultaneously maximize y_1 and y_4 and minimize y_2 and y_3. The design settings in the original and coded variables and the multiresponse data are given in Table 3.1.

Before we can find the simultaneous optima of the four responses, we need to find the 'best' multiresponse model for the data. To do so, we can use the PRESS and the Cp statistics for seemingly unrelated regression models derived by Sparks (1987). As an additional tool we can also use a measure of goodness of fit for seemingly unrelated regressions given by McElroy (1977), namely, McElroy's R^2. Detailed discussions of the PRESS and Cp criteria for multiresponse models and McElroy's R^2 are given in Appendix B, Sections B.1 and B.2 respectively. Because of the large number of subset models that we need to consider here, we will follow the recommendation of Sparks (1987) for reducing the number of subset models considered. First, by using the R^2, adjusted R^2, and the Cp criteria, we find all the 'good' subset models for y_i for $i = 1, 2, 3, 4$. For each y_i, $i = 1, 2, 3, 4$, we start with the model where all the linear

Table 3 1 Experimental Design and Response Values (example 1)

Original Input Variables			Coded Input Variables			Responses			
X_1	X_2	X_3	x_1	x_2	x_3	y_1	y_2	y_3	y_4
26 0	2 8	18 0	1 000	1 000	1 000	1 83	29 31	29 50	50 36
40 0	2 8	18 0	1 000	-1 000	-1 000	1 73	39 32	19 40	48 16
26 0	8 2	18 0	1 000	1 000	1 000	1 85	25 16	25 70	50 72
40 0	8 2	18 0	1 000	1 000	-1 000	1 67	40 84	27 10	46 69
26 0	2 8	27 0	1 000	1 000	1 000	1 86	29 82	21 40	50 09
40 0	2 8	27 0	1 000	-1 000	1 000	1 77	32 20	24 00	50 61
26 0	8 2	27 0	1 000	1 000	1 000	1 88	22 01	19 60	50 36
40 0	8 2	27 0	1 000	1 000	1 000	1 66	40 02	25 10	50 42
21 2	5 5	22 5	1 682	0 000	0 000	1 81	33 00	24 20	29 31
44 8	5 5	22 5	1 682	0 000	0 000	1 37	51 59	30 60	50 67
33 0	1 0	22 5	0 000	-1 682	0 000	1 85	20 35	20 90	48 75
33 0	10 0	22 5	0 000	1 682	0 000	1 92	20 53	18 90	52 70
33 0	5 5	14 9	0 000	0 000	-1 682	1 88	23 85	23 00	50 19
33 0	5 5	30 1	0 000	0 000	1 682	1 90	20 16	21 20	50 86
33 0	5 5	22 5	0 000	0 000	0 000	1 89	21 72	18 50	50 84
33 0	5 5	22 5	0 000	0 000	0 000	1 88	21 21	18 60	50 93
33 0	5 5	22 5	0 000	0 000	0 000	1 87	21 55	16 80	50 98

Source Khuri and Cornell (1996 pp 295 296)

terms x_1, x_2 and x_3 are included, then we increase the number of terms in the model by adding the second-degree (pure and cross product) terms one two or three at a time and then stop when all the terms in the model form the complete second-degree model Under the Cp criterion a subset model is considered 'good' if its Cp-value is close to p, the number of parameters in the subset model being considered Using R^2 and adjusted R^2 we say that a subset model is 'good' if the R^2 and adjusted R^2 values are close to 1 We consider all of these criteria to determine the 'best' models for each of the four responses of interest All these quantities were obtained by using the procedure PROC REG of SAS The models that are chosen for the four responses along with their R^2, adjusted R^2, and Cp values are given in Tables 3 2 – 3 5

Second, we examine all possible multiresponse models of the form given in equation (3 2 4) by including in its ith equation a 'good' subset model for y_i, $i = 1, 2, 3, 4$ This means that we consider all possible combinations of all the models for y_1, y_2, y_3,

Table 3 2 The 'Best' Models for y_1

Model No	R^2	Adj R^2	Cp	Variables in Model
3.4 1	0 945632	0 875730	10 0000	x_1 x_2 x_3, x_1^2 x_2^2 x_2^2 x_1x_2, x_1x_2, x_2x_3
3 4 2	0 944567	0 889133	8 1371	$x_1, x_2, x_3, x_1^2, x_2^2$, x_2^2, x_1x_2 x_1x_2
3 4 3	0 925385	0 880617	6 6068	x_1 x_2, x_3, x_1^2, x_2^2, x_1x_3
3 4 4	0 925393	0 880629	6 6057	x_1, x_2, x_3 x_1^2 x_2^2, x_3^2

Table 3 3 The Best Models for y_2

Model No	R^2	Adj R^2	Cp	Variables in Model
3.4 5	0 959286	0 906939	10 0000	$x_1, x_2, x_3, x_1^2, x_2^2$, $x_3^2, x_1x_2, x_1x_2, x_2x_3$
3 4 6	0 956786	0 913572	8 4298	x_1 x_2, x_2, x_1^2, x_2^2, x_3^2, x_1x_2, x_2x_3
3 4 7	0 948720	0.908836	7 8165	x_1, x_2, x_3, x_1^2 x_2^2 x_1x_2, x_2x_3

Table 3.4 The Best Models for y_3

Model No.	R^2	Adj R^2	Cp	Variables in Model
3 4.8	0 879106	0 723672	10 0000	x_1, x_2 x_3 x_1^2 x_2^2, $x_3^2, x_1x_2, x_1x_2, x_2x_2$
3 4 9	0.868857	0 737714	8 5935	$x_1, x_2, x_3, x_1^2, x_2^2$, x_3^2, x_1x_2, x_1x_3

Table 3 5 The Best Models for y_4

Model No	R^2	Adj R^2	Cp	Variables in Model
3 4 10	0 575700	0 030172	10 0000	x_1, x_2 x_2 x_1^2 x_2^2, $x_3^2, x_1x_2, x_1x_3, x_2x_2$
3 4 11	0 564945	0 129889	8 1774	x_1, x_2, x_2, x_1^2 x_2^2, x_1x_2 x_1x_2, x_1x_2
3 4 12	0 567180	0 134360	8 1400	x_1, x_2, x_3, x_1^2 x_2^2, x_1x_2 x_1x_3, x_2x_3

and y_4 given in Tables 3.2 - 3.5. For example, using Tables 3.2 - 3.5 one possible multiresponse model is formed by using model (3.4.1) for y_1, model (3.4.5) for y_2, model (3.4.8) for y_3, and model (3.4.10) for y_4. The next possible multiresponse model is given by the combination of models (3.4.1), (3.4.5), (3.4.8), and (3.4.11) then the next set is given by models (3.4.1) (3.4.5), (3.4.8), and (3.4.12), and so on until we get the last set given by models (3.4.4), (3.4.7), (3.4.9), and (3.4.12). If we do this we should get a total of $4 \times 3 \times 2 \times 3 = 72$ possible multiresponse models. For each resulting multirespoose model Sparks' PRESS statistic and McElroy's R^2 are then computed. The results for the 72 multiresponse models are given in Table 3.6. The Cp statistic was not computed for the resulting multiresponse models because according to Sparks (1987), the Cp criterion is only useful if $(n - \sum_{i=1}^{r} q_i)$ is fairly large, that is, greater than 25. In this case, the largest value that this quantity will have is $17 - 8 = 9$, which is quite small.

Finally, from Table 3.6 we choose the 'best' multiresponse model, that is, the model with the smallest PRESS statistic value and R^2 value close to 1. We select model number 48 in Table 3.6 because it clearly has the smallest PRESS statistic value and it also has an R^2 value quite close to that of the highest value associated with model number 1 in Table 3.6. We can then use model (3.4.4) for y_1, model (3.4.7) for y_2, model (3.4.9) for y_3, and model (3.4.11) for y_4.

Next, we need to determine the generalized least squares estimate $\boldsymbol{\gamma}$ for the multiresponse model formed by using models (3.4.4) (3.4.7), (3.4.9), and (3.4.11). The S matrix is given in Table 3.7. Using these values the matrix $\boldsymbol{\gamma}_e$ is computed and the resulting values for $\boldsymbol{\gamma}_{1e}$, $\boldsymbol{\gamma}_{2e}$, $\boldsymbol{\gamma}_{3e}$, and $\boldsymbol{\gamma}_{4e}$ are as follows:

$$\boldsymbol{\gamma}'_{1e} = (1.8766, -0.0974, -0.0009, 0.0091 \;\; -0.1020, 0.0044 \;\; 0.0037)',$$

$$\boldsymbol{\gamma}'_{2e} = (22.6128 \;\; 5.6609 \;\; -0.1719 \;\; -1.2268 \;\; 7.8551 \;\; 0.0663 \;\; 2.6920 \;\; 0.1752)'$$

$$\boldsymbol{\gamma}'_{3e} = (17.9782 \;\; 0.7442 \;\; -0.0120 \;\; -1.0710 \;\; 3.4414 \;\; 0.7004 \;\; 1.6328 \;\; 1.3020 \;\; 1.9716)'$$

Table 3.6 Multiresponse Models Formed by Using Models 3.4.18 to 3.4.29

Model No	Model for y_1	Model for y_2	Model for y_3	Model for y_4	Sparks' PRESS	Mc Elroy's R^2
1	3 4 1	3 4 5	3 4 8	3 4 10	518 48588	0 9676
2	3 4 2	3 4 5	3 4 8	3 4 10	502 94491	0 9673
3	3 4 3	3 4 5	3 4 8	3 4 10	459 31075	0 9630
4	3 4 4	3 4 5	3 4 8	3 4 10	454 30013	0 9630
5	3 4 1	3 4 6	3 4 8	3 4 10	499 60745	0 9671
6	3 4 2	3 4 6	3 4 8	3 4 10	469 84995	0 9668
7	3 4 3	3 4 6	3 4 8	3 4 10	420 97975	0 9625
8	3 4 4	3 4 6	3 4 8	3 4 10	413 75768	0 9625
9	3 4 1	3 4 7	3 4 8	3 4 10	471 52624	0 9655
10	3 4 2	3 4 7	3 4 8	3 4 10	444 00034	0 9652
11	3 4 3	3 4 7	3 4 8	3 4 10	400 27781	0 9605
12	3 4 4	3 4 7	3 4 8	3 4 10	391 39503	0 9605
13	3 4 1	3 4 5	3 4 9	3 4 10	482 99554	0 9669
14	3 4 2	3 4 5	3 4 9	3 4 10	470 64259	0 9654
15	3 4 3	3 4 5	3 4 9	3 4 10	424 42594	0 9613
16	3 4 4	3 4 5	3 4 9	3 4 10	420 11237	0 9613
17	3 4 1	3 4 6	3 4 9	3 4 10	462 8092	0 9664
18	3 4 2	3 4 6	3 4 9	3 4 10	435 84031	0 9649
19	3 4 3	3 4 6	3 4 9	3 4 10	385 72738	0 9607
20	3 4 4	3 4 6	3 4 9	3 4 10	379 26656	0 9607
21	3 4 1	3 4 7	3 4 9	3 4 10	436 06774	0 9648
22	3 4 2	3 4 7	3 4 8	3 4 10	411 01371	0 9633
23	3 4 3	3 4 7	3 4 9	3 4 10	366 95596	0 9588
24	3 4 4	3 4 7	3 4 9	3 4 10	358 10788	0 9588
25	3 4 1	3 4 5	3 4 8	3 4 11	510 76404	0 9674
26	3 4 2	3 4 5	3 4 8	3 4 11	495 54594	0 9671
27	3 4 3	3 4 5	3 4 8	3 4 11	452 40365	0 9628
28	3 4 4	3 4 5	3 4 8	3 4 11	447 38154	0 9628
29	3 4 1	3 4 6	3 4 8	3 4 11	490 57768	0 9669
30	3 4 2	3 4 6	3 4 8	3 4 11	461 26070	0 9666
31	3 4 3	3 4 6	3 4 8	3 4 11	413 04998	0 9622
32	3 4 4	3 4 6	3 4 8	3 4 11	405 63593	0 9622
33	3 4 1	3 4 7	3 4 8	3 4 11	466 19212	0 9654
34	3 4 2	3 4 7	3 4 8	3 4 11	438 90534	0 9651
35	3 4 3	3 4 7	3 4 8	3 4 11	395 15469	0 9604
36	3 4 4	3 4 7	3 4 8	3 4 11	386 15158	0 9604

Table 3 6 (continued)

Model No	Model for y_1	Model for y_2	Model for y_3	Model for y_4	Sparks PRESS	Mc Elroy's R^2
37	3 4 1	3 4 5	3 4 9	3 4 11	475 75777	0 9667
38	3 4 2	3 4 5	3 4 9	3 4 11	462 93734	0 9652
39	3 4 3	3 4 5	3 4 9	3 4 11	417 65202	0 9610
40	3 4 4	3 4 5	3 4 9	3 4 11	413 39668	0 9611
41	3 4 1	3 4 6	3 4 9	3 4 11	454 21890	0 9662
42	3 4 2	3 4 6	3 4 9	3 4 11	426 86098	0 9647
43	3 4 3	3 4 6	3 4 9	3 4 11	377 83619	0 9604
44	3 4 4	3 4 6	3 4 9	3 4 11	371 26193	0 9605
45	3 4 1	3 4 7	3 4 8	3 4 11	431 00926	0 9647
46	3 4 2	3 4 7	3 4 9	3 4 11	405 61274	0 9632
47	3 4 3	3 4 7	3 4 9	3 4 11	361 63956	0 9587
48	3 4 4	3 4 7	3 4 9	3 4 11	352 81731	0 9587
49	3 4 1	3 4 5	3 4 8	3 4 12	510 03862	0 9675
50	3 4 2	3 4 5	3 4 8	3 4 12	494 91959	0 9671
51	3 4 3	3 4 5	3 4 8	3 4 12	452 12913	0 9626
52	3 4 4	3 4 5	3 4 8	3 4 12	447 48627	0 9629
53	3 4 1	3 4 6	3 4 8	3 4 12	489 84376	0 9656
54	3 4 2	3 4 6	3 4 8	3 4 12	460 62969	0 9652
55	3 4 3	3 4 6	3 4 8	3 4 12	412 05971	0 9605
56	3 4 4	3 4 6	3 4 8	3 4 12	405 77815	0 9605
57	3 4 1	3 4 7	3 4 8	3 4 12	475 29172	0 9655
58	3 4 2	3 4 7	3 4 8	3 4 12	447 66382	0 9652
59	3 4 3	3 4 7	3 4 8	3 4 12	401 82939	0 9605
60	3 4 4	3 4 7	3 4 8	3 4 12	394 59055	0 9605
61	3 4 1	3 4 5	3 4 9	3 4 12	474 79406	0 9668
62	3 4 2	3 4 5	3 4 9	3 4 12	462 07602	0 9652
63	3 4 3	3 4 5	3 4 9	3 4 12	417 20696	0 9608
64	3 4 4	3 4 5	3 4 9	3 4 12	413 19236	0 9611
65	3 4 1	3 4 6	3 4 9	3 4 12	453 28532	0 9663
66	3 4 2	3 4 6	3 4 9	3 4 12	426 02696	0 9647
67	3 4 3	3 4 6	3 4 9	3 4 12	376 73766	0 9603
68	3 4 4	3 4 6	3 4 9	3 4 12	371 13377	0 9605
69	3 4 1	3 4 7	3 4 9	3 4 12	439 57150	0 9648
70	3 4 2	3 4 7	3 4 9	3 4 12	414 29722	0 9633
71	3 4 3	3 4 7	3 4 9	3 4 12	367 52471	0 9588
72	3 4 4	3 4 7	3 4 9	3 4 12	360 61816	0 9588

Table 3.7. The S matrix

0.001288	0.001632	0.024424	-0.067885
0.001288	0.001632	0.024424	-0.067885
0.024424	0.397570	1.990787	2.789313
0.067885	1.978386	2.789313	11.090586

$$\gamma'_{4c} = (51.9454\ 2.4364\ 0.6307\ 0.2692\ -3.43691\ -0.0358\ -1.2953\ 0.6010, 0.1260)'$$

Thus the models for the four responses are given by

$$\begin{aligned}
y_{1c} &= 1.8765 - 0.0974x_1 - 0.0009x_2 + 0.0091x_3 - 0.1020x_1^2 + \\
&\quad 0.0044x_2^2 + 0.0037x_3^2 \\
y_{2c} &= 22.6128 + 5.6609x_1 - 0.1719x_2 - 1.2268x_3 + 7.8551x_1^2 + \\
&\quad 0.0663x_2^2 + 2.6920x_1x_2 + 0.1752x_2x_3 \\
y_{3c} &= 17.9782 + 0.7442x_1 - 0.0120x_2 - 1.0710x_3 + 3.4414x_1^2 + \\
&\quad 0.7004x_2^2 + 1.6328x_3^2 + 1.3020x_1x_2 + 1.9716x_1x_3 \\
y_{3c} &= 51.9454 + 2.4364x_1 + 0.6307x_2 + 0.2692x_3 - 3.43691x_1^2 - \\
&\quad 0.0358x_2^2 - 1.2953x_1x_2 + 0.6010x_1x_3 + 0.1260x_2x_3 \qquad (3.4.13)
\end{aligned}$$

Using the γ_{ic} values, we can now find the individual and simultaneous optima of the four responses. The individual optima were obtained by using a FORTRAN program written by Conlon (1992) which applies the optimization algorithm of Price (1977). A description of this algorithm is given in Appendix B, Section B.3. The individual optima and their locations are given in Table 3.8. Note that the experimental region is given by $\sum_{i=1}^{3} x_i^2 \leq 3$. The distance measure ρ in equation (3.3.1) is directly minimized using the procedure PROC IML of SAS where 16 000 values of ρ were computed. The minimum value of ρ, the simultaneous optima of the four responses, and their location are given in Table 3.9. Note that the simultaneous optima of y_1, y_2, y_3, and y_4 are close to their corresponding individual maxima/minima.

Table 3.8. The Individual Optima of the Four Responses

Response	Individual Optima	Location
$\hat{y}_{1e}$	$\phi_1 = 1\ 9271$	$(-0.4428, -0\ 0113\ 1\ 6690)'$
$\hat{y}_{2e}$	$\phi_2 = 18\ 8366$	$(-0\ 5524\ 1\ 3082\ 0\ 9917)'$
$\hat{y}_{3e}$	$\phi_3 = 17\ 5833$	$(-0\ 3096\ 0\ 3028\ 0\ 5161)'$
$\hat{y}_{4e}$	$\phi_4 = 53\ 4116$	$(0\ 5488\ 0\ 5427, 1\ 5503)'$

Table 3.9. The Simultaneous Optima of the Four Responses

Response	Simultaneous Optima
$\hat{y}_{1e}$	1.9166
$\hat{y}_{2e}$	18 9631
$\hat{y}_{3e}$	18 4019
$\hat{y}_{4e}$	51 9021
Location	$\boldsymbol{x} = (-0\ 4917, 1\ 4904, 0\ 7079)'$
min μ	1 8514

It should also be noted that the minimax approach based on Theorem 3 3 1 can also be applied in this example.

The optimization procedures using the desirability function aproach by Derringer and Suich (1980) and the nonlinear optimization procedure using the GRG algorithm were also used in this example. These two procedures were discussed in Chapter 2, Section 2 4. The results for the desirability function approach are given in Table 3 10 and the results for the GRG approach are given in Table 3 11. The simultaneous optima and the desirability values in Table 3 10 were computed using the Design Expert software (see Design Expert Version 4 0 User's Guide for details on how to use the program) and the computer output is given in Appendix C. When using the desirability function approach, desirability values close to 1 are desired. The overall desirability D^* given in Table 3 10 is the geometric mean of the individual desirability values, d_i^*, defined in equations (2 4 2) and (2 4 3). In this example, Design Expert yielded a satisfactory value of 0 7018 for D^*. The results for the GRG algorithm approach are computed using the solver tool available in the software Excel. For details on how to use the solver tool of Microsoft Excel, see Dodge et al. (1995).

Table 3.10. Results for the Desirability Function Approach

Response	Simultaneous Optima	d_i^*
y_{1e}	1.92	0.52
y_{2e}	19.13	0.94
y_{3e}	17.57	0.78
y_{4e}	48.98	0.64
Location	$x = (-0.6180, 1.000, 0.9999)'$	
D^*	0.7018	

Table 3.11. Results for the GRG Approach

Primary Response	Optimum for Primary Response	Location
y_{1e}	1.9177	$(-0.5243, 1.4080, 0.8618)'$
y_{2e}	20.7467	$(-0.2335, 0.4916, 0.5218)'$
y_{3e}	18.4500	$(-0.4578, 1.3580, 0.9728)'$
y_{4e}	51.2443	$(-0.5842, 1.3308, 0.9422)'$

The results given in Table 3.11 are the optima for each of the four responses, that is, each of the four responses was considered to be the primary response and the optimum value was obtained using the constraints that the other three responses are equal to their respective individual optima. For example, if y_1 is considered to be the primary response, then the maximum of y_1 is equal to 1.9177. This optimum value is obtained subject to the constraints that y_{2e}, y_{3e}, and y_{4e} given in equation (3.4.13) are equal to 18.8366, 17.5833, 53.4116, respectively. The constraint $\sum_{i=1}^{3} x_i^2 \leq 3$ was also used. The same procedure was followed for the other three cases, with the constraints changed, depending on which variable is considered as the primary response.

The optima using the generalized distance approach, the desirability function approach and the GRG algorithm approach, given in Tables 3.9, 3.10 and 3.11 respectively, are quite close to each other. For y_1, the maximum value using the generalized distance approach, 1.9166, is very close to that obtained using the desirability function approach, that is, 1.9177. These two values are also very close to 1.92, the value obtained using the GRG approach, with y_1 considered as the primary response. The

Table 3.12 The Original and Coded Levels of the Input Variables

	Coded Levels				
Variable	2	1	0	1	2
X_1	65 000	70 000	75 000	80 000	85 000
X_2	4 000	5 000	6 000	7 000	8 000
X_3	0 025	0 075	0 175	0 275	0 375
X_4	0 000	0 0125	0 025	0375	0 050
X_5	0 000	0 050	0 100	0.150	0 200

Source: Khuri and Cornell (1996 p 278)

minimum values for y_2 obtained using these three approaches are also not that different The same can be said about the minimum values of y_3 and the maximum values of y_4 One should note however, the differences in the way these optima were obtained As mentioned in Chapter 2, Section 2 4 the desirability function approach and the GRG algorithm approach do not take into consideration the correlations among the responses and the magnitudes of their variances These are important considerations in multiresponse optimization (see Khuri and Cornell (1996), p 285) The generalized distance approach takes these into account Hence it is recommended that the results for the generalized distance approach be used

Example 2 Richert et al (1974) investigated the effects of heating temperature (x_1), pH level (x_2), redox potential (x_3) sodium oxalate (x_4), and sodium lauryl sulfate (x_5) on foaming properties of whey protein concentrates The responses were whipping time (y_1), maximum overrun (y_2) and soluble protein (y_3) A central composite design consisting of a 2^{5-1} fractional factorial design ten axial points, and five center point replications was used The results of this study were utilized to find the simultaneous maxima of all three responses. The original and coded levels of the five input variables are given in Table 3 12 and the data set is given in Table 3 13

To determine the 'best' multiresponse model in terms of the PRESS criterion and R^2 we use the same procedure as described in Example 1 The subset models given in Tables 3 14, 3 15, and 3 16 were chosen to be the 'best' (based on the Cp-criterion,

Table 3 13 Experimental Design and Response Values (Example 2)

x_1	x_2	x_3	x_4	x_5	y_1	y_2	y_3
0	0	0	0	0	3 5	1179	104 0
0	0	0	0	0	3 5	1183	107 0
0	0	0	0	0	4 0	1120	104 0
0	0	0	0	0	3 5	1180	101 0
0	0	0	0	0	3.0	1195	103 0
-1	-1	-1	-1	1	4 75	1082	81 4
1	1	-1	-1	-1	4 0	824	69 6
1	1	-1	1	1	5.0	953	105 0
1	1	1	1	1	9 5	759	81 2
-1	1	1	1	1	4 0	1163	80 8
1	1	1	1	1	5 0	839	76.3
1	1	1	-1	1	3 0	1343	103 0
1	1	1	-1	-1	7 0	736	76 9
1	-1	-1	1	-1	5 25	1027	87 2
1	1	-1	1	1	5 0	836	74 0
-1	1	-1	1	1	3 0	1272	98 5
1	1	-1	1	1	6 5	825	94 1
1	1	1	1	1	3 25	1363	95 9
1	1	1	1	1	5 0	855	76.8
1	1	1	1	-1	2 75	1284	100 0
1	1	1	1	1	5 0	851	104 0
2	0	0	0	0	3 75	1263	100 0
2	0	0	0	0	11 0	651	50 5
0	2	0	0	0	4 5	1217	71 2
0	2	0	0	0	4 0	982	101 0
0	0	-2	0	0	5 0	884	85 8
0	0	2	0	0	3 75	1147	103 0
0	0	0	-2	0	3 75	1081	104 0
0	0	0	2	0	4 75	1036	89 4
0	0	0	0	-2	4 0	1213	105 0
0	0	0	0	2	3 5	1103	113 0

Source Khuri and Cornell (1996, p 279)

Table 3 14 The Best Models for y_3

Model No.	R^2	Adj R^2	Cp	Variables in Model
3 4 14	0 91826	0 75478	21 000	x_1 x_2 x_3, x_4, x_5, x_1^2 x_2^2, x_3^2, x_4^2, x_5^2, x_1x_2, x_1x_3, x_1x_4, x_1x_5, x_2x_3, x_2x_4, x_3x_4, x_3x_5, x_4x_5
3 4 15	0 91258	0 76157	19 695	x_1, x_2, x_3, x_4, x_5, x_1^2, x_2^2, x_3^2, x_5^2, x_1x_2, x_1x_3, x_1x_4, x_1x_5, x_2x_3, x_2x_4, x_2x_5, x_3x_4, x_3x_5, x_4x_5
3 4 16	0 91258	0 76157	19 695	x_1 x_2, x_3, x_4, x_5, x_1^2, x_2^2, x_3^2, x_4^2, x_5^2, x_1x_2 x_1x_3, x_1x_4, x_1x_5, x_2x_3, x_2x_4, x_2x_5, x_3x_4, x_3x_5, x_4x_5
3 4 17	0 83027	0 75752	9 765	x_1, x_2 x_3, x_4, x_5, x_1^2, x_1x_2, x_2x_3, x_4x_5

R^2, and adjusted R^2 values) subset models for each of the three responses y_1, y_2, and y_3, respectively

All possible multiresponse models are then formed from the models given in Tables 3 14 to 3 16 and the PRESS statistic and R^2 values are given in Table 3 17 We choose model number 64 in Table 3 17 as the 'best multiresponse model because it has the smallest PRESS statistic value and its R^2 value, even though it is not the highest is quite close to the highest value associated with model number 1 For the individual responses we can then consider model (3 4 17) for y_1, model (3 4 21) for y_2, and model (3 4 25) for y_3

Table 3.15 The Best' Models for y_2

Model No	R^2	Adj R^2	Cp	Variables in Model
3 4 18	0 93827	0 81482	21 000	x_1, x_2, x_3, x_4, x_5 $x_1^2, x_2^2, x_3^2,$ $x_4^2, x_5^2, x_1x_2,$ $x_1x_3, x_1x_4, x_1x_5,$ x_2x_3, x_2x_4, x_2x_5 x_3x_4, x_3x_5, x_4x_5
3 4 19	0 93439	0 82106	19 630	$x_1, x_2, x_3, x_4, x_5,$ $x_1^2, x_2^2, x_3^2,$ $x_4^2, x_5^2, x_1x_2,$ $x_1x_3, x_1x_4, x_1x_5,$ x_2x_3, x_2x_5, x_3x_4 x_3x_5, x_4x_5
3 4 20	0 87315	0 81879	9 550	$x_1, x_2, x_3, x_4, x_5,$ $x_1^2, x_3^2,$ x_1x_2, x_1x_3
3 4 21	0 86449	0 81522	8 953	$x_1, x_2, x_3, x_4, x_5,$ x_1^2, x_3^2, x_1x_3

Table 3.16 The Best Models for y_3

Model No.	R^2	Adj R^2	Cp	Variables in Model
3 4 22	0 91340	0 74019	21 000	$x_1 \ x_2 \ x_3, x_4 \ x_5$ $x_1^2, x_2^2, x_3^2,$ $x_4^2, x_5^2, x_1x_2,$ x_1x_3, x_1x_4, x_1x_5 $x_2x_3, x_2x_4, x_2x_5,$ x_3x_4, x_3x_5, x_4x_5
3 4 23	0 90819	0 74960	19 601	$x_1, x_2, x_3, x_4, x_5,$ $x_1^2, x_2^2, x_3^2,$ $x_4^2, x_5^2, x_1x_2,$ x_1x_3, x_1x_4, x_1x_5 x_2x_1, x_2x_4, x_2x_3 x_3x_5, x_4x_5
3 4 24	0 90905	0 75195	19 502	$x_1, x_2, x_3, x_4, x_5,$ $x_1^2, x_2^2, x_3^2,$ $x_4^2 \ x_1x_2, x_1x_3,$ x_1x_4, x_1x_5, x_2x_3 $x_2x_4, x_2x_5,$ x_3x_5, x_4x
3 4 25	0 81255	0 74439	8 644	$x_1, x_2, x_3, x_4, x_5,$ x_1^2, x_2^2, x_4^2

Table 3 17 Multiresponse Models Formed by Using Models 3 4 14 to 3 4 25

Model No.	Model for y_1	Model for y_2	Model for y_3	PRESS	R^2
1	3 4 14	3 4 18	3 4 22	2154 3046	0 9445
2	3 4 15	3 4 18	3 4 22	2027 3294	0 9433
3	3 4 16	3 4 18	3 4 22	2025 1836	0 9433
4	3 4 17	3 4 18	3 4 22	1680 8867	0 9246
5	3 4 14	3 4 19	3 4 22	1947 4140	0 9250
6	3 4 15	3 4 19	3 4 22	1803 8604	0 9421
7	3 4 16	3 4 19	3 4 22	1801 7832	0 9421
8	3 4 17	3 4 19	3 4 22	1433 4468	0 9220
9	3 4 14	3 4 20	3 4 22	1616 1234	0 9225
10	3 4 15	3 4 20	3 4 22	1513 9965	0 9223
11	3 4 16	3 4 20	3 4 22	1478 3295	0 9228
12	3 4 17	3 4 20	3 4 22	1037 0494	0 8996
13	3 4 14	3 4 21	3 4 22	1610 0459	0 9225
14	3 4 15	3 4 21	3 4 22	1506 0677	0 9197
15	3 4 16	3 4 21	3 4 22	1471 9379	0 9202
16	3 4 17	3 4 21	3 4 22	1021 4802	0 8970
17	3 4 14	3 4 18	3 4 23	1885 5151	0 9430
18	3 4 15	3 4 18	3 4 23	1751 3511	0 9417
19	3 4 16	3 4 18	3 4 23	1749 9175	0 9417
20	3 4 17	3 4 18	3 4 23	1390 0528	0 9206
21	3 4 14	3 4 19	3 4 23	1604 4202	0 9223
22	3 4 15	3 4 19	3 4 23	1455 3973	0 9405
23	3 4 16	3 4 19	3 4 23	1454 0606	0 9405
24	3 4 17	3 4 19	3 4 23	1068 8917	0 9178
25	3 4 14	3 4 20	3 4 23	1259 5952	0 9196
26	3 4 15	3 4 20	3 4 23	1149 4299	0 9195
27	3 4 16	3 4 20	3 4 23	1116 0802	0 9200
28	3 4 17	3 4 20	3 4 23	940 1738	0 8930
29	3 4 14	3 4 21	3 4 23	1257 3202	0 9196
30	3 4 15	3 4 21	3 4 23	1145 5039	0 9187
31	3 4 16	3 4 21	3 4 23	1113 5910	0 9172
32	3 4 17	3 4 21	3 4 23	630 3612	0 8903

Table 3 17 (continued)

Model No	Model for y_1	Model for y_2	Model for y_3	PRESS	R^2
33	3 4 14	3 4 18	3 4 24	2050 8224	0 9433
34	3 4 15	3 4 18	3 4 24	1913 5294	0 9417
35	3 4 16	3 4 18	3 4 24	1912 5955	0 9417
36	3 4 17	3 4 18	3 4 24	1583 4237	0 9213
37	3 4 14	3 4 19	3 4 24	1812 6690	0 9225
38	3 4 15	3 4 19	3 4 24	1660 0629	0 9405
39	3 4 16	3 4 19	3 4 24	1659 2509	0 9406
40	3 4 17	3 4 19	3 4 24	1307 8545	0 9186
41	3 4 14	3 4 20	3 4 24	1489 7999	0 9198
42	3 4 15	3 4 20	3 4 24	1382 8109	0 9193
43	3 4 16	3 4 20	3 4 24	1348 1879	0 9198
44	3 4 17	3 4 20	3 4 24	911 1950	0 8933
45	3 4 14	3 4 21	3 4 24	1485 1354	0 9198
46	3 4 15	3 4 21	3 4 24	1375 5986	0 9166
47	3 4 16	3 4 21	3 4 24	1342 6031	0 9171
48	3 4 17	3 4 21	3 4 24	897 2984	0 8906
49	3 4 14	3 4 18	3 4 25	1822 0860	0 9275
50	3 4 15	3 4 18	3 4 25	1506 4869	0 9258
51	3 4 16	3 4 18	3 4 25	1507 8704	0 9258
52	3 4 17	3 4 18	3 4 25	1037 5894	0 8976
53	3 4 14	3 4 19	3 4 25	1269 4878	0 8993
54	3 4 15	3 4 19	3 4 25	1143 1515	0 9242
55	3 4 16	3 4 19	3 4 25	1144 8639	0 9242
56	3 4 17	3 4 19	3 4 25	588 5999	0 8917
57	3 4 14	3 4 20	3 4 25	950 4228	0 8956
58	3 4 15	3 4 20	3 4 25	847 2803	0 8956
59	3 4 16	3 4 20	3 4 25	827 7270	0 8963
60	3 4 17	3 4 20	3 4 25	256 2547	0 8554
61	3 4 14	3 4 21	3 4 25	951 0136	0 8956
62	3 4 15	3 4 21	3 4 25	845 1851	0 8920
63	3 4 16	3 4 21	3 4 25	826 8527	0 8926
64	3 4 17	3 4 21	3 4 25	249 5952	0 8517

Table 3.18. The S matrix

0 53375	12 65426	−2 45327
12 65426	5048 52800	23 15932
−2 45327	23 15932	38 71550

The $\boldsymbol{S}$ matrix is determined by using the individual models and is given in Table 3 18 From these values $\hat{\boldsymbol{\gamma}}$ is computed yielding

$$\begin{aligned}
\boldsymbol{\gamma}'_{1e} &= (3\,9474, 1\,2708\ \ 0\,1875\ \ -0\,4375\ \ -0\,1875, -0\,0833,\\
&\qquad 0\,8805\ \ 0\,7191, -0\,3361, -0.3277)'.\\
\boldsymbol{\gamma}'_{2e} &= (1115.396\ \ -176\,0833\ \ -18\,1667\ \ 57\,5833, 21\,8333, 19\,0833,\\
&\qquad -50\,4371\ \ -37\,5328, -44\,5806)'\\
\boldsymbol{\gamma}'_{3e} &= (101\,8229, -8\,2458, 7\,5125\ \ 2\,3792\ \ 1\,1292\ \ 1\,6625\\
&\qquad -7\,2013, -4\,1430\ \ -1\,5686)'
\end{aligned}$$

Using the $\hat{\boldsymbol{\gamma}}_{ie}$ values we have

$$\begin{aligned}
y_{1e} &= 3\,9474 + 1\,2708x_1 + 0\,1875x_2 - 0\,4375x_3 - 0\,1875x_4 - 0\,0833x_5 +\\
&\quad 0\,8805x_1^2 + 0\,7191x_1x_2 - 0\,3361x_2x_3 - 0.3277x_4x_5\\
y_{2e} &= 1115\,396 - 176\,0833x_1 - 18\,1667x_2 + 57\,5833x_3 + 21\,8333x_4 + 19\,0833x_5 -\\
&\quad 50\,4371x_1^2 - 37\,5328x_3^2 - 44.5806x_1x_3\\
y_{3e} &= 101\,8229 - 8\,2458x_1 + 7\,5125x_2 + 2.3792x_3 + 1\,1292x_4 + 1\,6625x_5 -\\
&\quad 7\,2013x_1^2 - 4\,1430x_2^2 - 1\,5686x_4^2 \qquad (3\,4\,26)
\end{aligned}$$

The individual maxima are again obtained using the FORTRAN program by Conlon (1992) (see Appendix B Section B 3) These maximum values and their locations are given in Table 3 19 The experimental region is given by $\sum_{i=1}^{5} x_i^2 \leq 5$ The minimum of the distance measure ρ given in equation (3 3 1) is found based on 26,000 values of ρ using the procedure PROC IML of SAS Its value and location are given in Table 3 20 The simultaneous maxima are also given in Table 3.20 These simultaneous maxima are lower than their corresponding individual maxima with the one for y_1 dropping from 11 8031 to 5 9897 and the one for y_3 dropping from 113 5343

Table 3 19 The Individual Maxima of the Three Responses

Response	Individual Maxima	Location
y_{1e}	$\phi_1 = 11\ 8031$	$(1\ 9991, 0\ 9248\ -0.3580\ -0\ 1378\ 0\ 0093)'$
y_{2e}	$\phi_2 = 1385\ 7395$	$(1\ 4292, 0\ 3501\ 1\ 1633\ 0\ 9862\ 0\ 1833)'$
y_{3e}	$\phi_3 = 113\ 5343$	$(-0\ 5185, 0\ 7689\ 1\ 6597, 0\ 2509\ 1\ 1498)'$

Table 3 20 The Simultaneous Maxima of the Three Responses

Response	Simultaneous Maxima
y_{1e}	5 9897
y_{2e}	1289 1471
y_{3e}	78 9504
Location $\boldsymbol{x} = (-1\ 9103, -1\ 0731\ -0\ 0287\ -0\ 0390\ 0\ 3628)'$	
max ρ	17 2141

to 78 9504 The maximax approach described in the previous section can also be used in this example.

CHAPTER 4

MULTIRESPONSE MODELS WITH FIXED BLOCK EFFECTS

4.1 Introduction

For the most part, analyses results available in the statistics literature concerning response surface models with block effects apply only to the single-response case. In experimental situations where several responses of interest are measured and where it is impossible to perform all the runs under homogeneous conditions, the analysis of the resulting multiresponse data must take into account heterogeneity of the experimental conditions. For example, a chemical engineer may choose to run an experiment using two separate machines (instead of only one) located in two different buildings. In this case, machine confounded with location should be included in the analysis of the experiment. This extraneous source of variation, that is, the effect of location (block effect), can influence the results of the analysis of the data, specifically, the estimation of the mean response over a given region of interest. Thus such block effects must be considered in any multiresponse analysis of the data.

In this chapter, the results given by Khuri (1994) concerning the effect of blocking on the estimation of a response surface are extended to the multiresponse case. Here we consider the estimation of the model parameters and of the mean response, the effect of blocking on the prediction variance, and present a condition for orthogonal blocking. These are all important considerations in the analysis of response surface

experiments with block effects The model assumptions and notation follow Khuri (1994) and these are given in the next section

4.2 Estimating the Parameters in Multiresponse Models with Fixed Block Effects

Khuri (1994) discussed the effect of blocking on the estimation of a response surface for the single-response case He considered the model

$$\boldsymbol{y}^* = \beta_0 \mathbf{1}_n + \boldsymbol{X}^* \boldsymbol{\beta}^* + \boldsymbol{Z}^* \boldsymbol{\delta}^* + \boldsymbol{\epsilon}^* \tag{4.2.1}$$

where $\boldsymbol{y}^*$ is the response vector which consists of n observations $\mathbf{1}_n$ is a vector of ones of order $n \times 1$ $\boldsymbol{X}^*$ is an $n \times p$ matrix whose uth row is $\boldsymbol{f}'(\boldsymbol{x}_u)$, where $\boldsymbol{x}_u$ is the vector of design settings at the uth experimental run $(u = 1, 2, \ldots, n)$ and $\boldsymbol{f}'(\boldsymbol{x}_u)$ is a vector of order $1 \times p$ whose elements consist of the x_{ul} terms $(l = 1\ 2 \ldots k)$ along with their powers and cross-products of these powers, $\boldsymbol{\beta}^*$ is a vector of unknown parameters $\boldsymbol{\delta}^* = (\delta_1\ \delta_2, \ldots \delta_b)'$ where δ_j denotes the effect of the jth block (j=1,2 ... b), and $\boldsymbol{Z}^*$ is a block-diagonal matrix of the form $\boldsymbol{Z} = diag(\mathbf{1}_{n_1}, \mathbf{1}_{n_2} \ldots, \mathbf{1}_{n_b})$, where n_j is the size of the jth block, that is, the blocks are of different sizes The random error vector $\boldsymbol{\epsilon}^*$ is assumed to have a zero mean and a variance-covariance matrix $\sigma^2 \boldsymbol{I}_n$

Model (4.2.1) can be written in a more compact form Since the columns of $\boldsymbol{Z}^*$ sum to $\mathbf{1}_n$, model (4.2.1) can be written as

$$\boldsymbol{y}^* = \boldsymbol{X}^* \boldsymbol{\beta}^* + \boldsymbol{Z}^* \boldsymbol{\tau}^* + \boldsymbol{\epsilon}^* \tag{4.2.2}$$

where $\boldsymbol{\tau}^* = \beta_0 \mathbf{1}_b + \boldsymbol{\delta}^*$ If the block effects are constrained to sum to zero then $\beta_0 = (\mathbf{1}/b) \sum_{j=1}^{b} \tau_j^*$ where τ_j^* is the jth element of $\boldsymbol{\tau}^*$ Model (4.2.2) can be written as

$$\boldsymbol{y}^* = \boldsymbol{W}^* \boldsymbol{\theta}^* + \boldsymbol{\epsilon}^* \tag{4.2.3}$$

where $\boldsymbol{W}^* = [\boldsymbol{X}^*\ \ \boldsymbol{Z}^*]$ and $\boldsymbol{\theta}^* = (\boldsymbol{\beta}^*\ \ \boldsymbol{\tau}^*)'$

We now extend the above results to the multiresponse case, that is obtain the multiresponse model and determine the estimates of the model s parameters. Suppose

there are r responses of interest and that all the reponses have the same numbes of observations n. The polynomial models for the different responses can have different degrees d_i. In some multiresponse experiments it is possible that the blocks used are not necessarily the same for all the responses. For example, a chemical engineer in a production plant wants to improve the operating conditions for the production of a certain polymer; that is, he/she wants to determine conditions on the input variables temperature, pressure, concentration of material A, and dye concentration that will optimize the responses given by mechanical strength and coloration of the polymer. Note that material A is used to affect the mechanical strength of the polymer and does not really affect coloration and that the dye does not really affect mechanical strength. The engineer found out that there are two batches of material A and three batches of dye available for the experiment. In this case, two different blocks may be used for the two responses, that is, the 2 batches of material A for mechanical strength and the 3 batches of the dye for coloration. For example, if a CCD for $k = 4$ factors is used to run this experiment, then for mechanical strength, the two blocks given by the factorial portion with center point replicates and the axial portion with center point replicates may be used. For coloration, two of the three blocks each consists of a half replicate of the 2^4 factorial points plus center point replicates and the third block consists of the axial points plus center point replicates. Thus in setting up the multiresponse model, we will consider the case in which the blocks are not necessarily the same for all the responses. From models (4.2.1) and (4.2.3), the ith single-response model is given by

$$\boldsymbol{y}_i = \beta_{0i}\mathbf{1}_n + \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{Z}_i\boldsymbol{\delta}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, r \tag{4.2.4}$$

or equivalently,

$$\boldsymbol{y}_i = \boldsymbol{W}_i\,\boldsymbol{\theta}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, r \tag{4.2.5}$$

where $\boldsymbol{W}_i = [\boldsymbol{X}_i \quad \boldsymbol{Z}_i]$ $\boldsymbol{Z}_i = \text{diag}(\mathbf{1}_{n_{i1}}, \mathbf{1}_{n_{i2}}, \ldots, \mathbf{1}_{n_{ib_i}})$ where n_{ij} is the size of the jth block for the ith response b_i is the number of blocks for the ith response, $\sum_{j=1}^{b_i} n_{ij} = n$, $i = 1, 2, \ldots, r$ $\boldsymbol{X}_i$ is of order $n \times p_i$ $\boldsymbol{\theta}_i = (\boldsymbol{\beta}_i^{'} \ \boldsymbol{\tau}_i^{'})^{'}$ $\boldsymbol{\beta}_i$ is a vector of unknown parameters; $\boldsymbol{\tau}_i = \beta_{0i}\mathbf{1}_{b_i} + \boldsymbol{\delta}_i$, and $\boldsymbol{\delta}_i = (\delta_{i1}, \delta_{i2}, \ldots, \delta_{ib_i})^{'}$, where δ_{ij} denotes the effect of the jth block in the ith response model, $j = 1, 2, \ldots, b_i$ $i = 1, 2, \ldots, r$ It is assumed that

$$E(\boldsymbol{\epsilon}_i) = \mathbf{0}, \ i = 1, 2, \ldots, r, \quad \text{and} \quad E(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_l^{'}) = \sigma_{il}\boldsymbol{I}_n, \ i, l = 1, 2, \ldots, r \tag{4.2.6}$$

The response models in (4.2.5) can be expressed as a single multiresponse model of the form

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{\theta} + \boldsymbol{\epsilon} \tag{4.2.7}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1^{'} \ \boldsymbol{y}_2^{'} \ \ldots \ \boldsymbol{y}_r^{'})^{'}$, $\boldsymbol{W}$ is a block diagonal matrix of the form $\boldsymbol{W} = \text{diag}(\boldsymbol{W}_1, \boldsymbol{W}_2, \ldots, \boldsymbol{W}_r)$ $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^{'}, \boldsymbol{\theta}_2^{'}, \ldots, \boldsymbol{\theta}_r^{'})^{'}$ and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^{'}, \boldsymbol{\epsilon}_2^{'}, \ldots, \boldsymbol{\epsilon}_r^{'})^{'}$ By (4.2.6), the variance-covariance matrix of $\boldsymbol{\epsilon}$ is $\boldsymbol{\Sigma} \otimes \boldsymbol{I}_n$, where $\boldsymbol{\Sigma} = (\sigma_{ij})$ is the variance-covariance matrix of the r responses.

The estimates of the parameters in model (4.2.7) can be obtained by using results from Srivastava and Giles (1987 p.12) The generalized least squares estimator of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{W}^{'}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{W})^{-1}\boldsymbol{W}^{'}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} \tag{4.2.8}$$

and its variance-covariance matrix is

$$Var(\hat{\boldsymbol{\theta}}) = [\boldsymbol{W}^{'}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{W}]^{-1} \tag{4.2.9}$$

Let $\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n = (\sigma^{ij}\boldsymbol{I}_n)$ Then using the result given by Zellner (1962 p.351, formula 2.5), the variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ can be written in the form

$$Var(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \sigma^{11}\boldsymbol{W}_1^{'}\boldsymbol{W}_1 & \sigma^{12}\boldsymbol{W}_1^{'}\boldsymbol{W}_2 & \cdots & \sigma^{1r}\boldsymbol{W}_1^{'}\boldsymbol{W}_r \\ \sigma^{21}\boldsymbol{W}_2^{'}\boldsymbol{W}_1 & \sigma^{22}\boldsymbol{W}_2^{'}\boldsymbol{W}_2 & \cdots & \sigma^{2r}\boldsymbol{W}_2^{'}\boldsymbol{W}_r \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{r1}\boldsymbol{W}_r^{'}\boldsymbol{W}_1 & \sigma^{r2}\boldsymbol{W}_r^{'}\boldsymbol{W}_2 & \cdots & \sigma^{rr}\boldsymbol{W}_r^{'}\boldsymbol{W}_r \end{pmatrix}^{-1} \tag{4.2.10}$$

In general, the variance-covariance matrix $\boldsymbol{\Sigma}$ is unknown. To obtain the estimates in (4.2.8) $\boldsymbol{\Sigma}$ must be estimated. To estimate $\boldsymbol{\Sigma}$ we can use the estimator given by Zellner (1962) (see also Srivastava and Giles (1987 pp 13-17)). This estimator was denoted in Chapter 3 by $\boldsymbol{S} = (s_{ij})$ where

$$s_{ij} = \frac{1}{n} \boldsymbol{y}_i' [\boldsymbol{I} - \boldsymbol{W}_i (\boldsymbol{W}_i' \boldsymbol{W}_i)^{-1} \boldsymbol{W}_i'] [\boldsymbol{I} - \boldsymbol{W}_j (\boldsymbol{W}_j' \boldsymbol{W}_j)^{-1} \boldsymbol{W}_j'] \boldsymbol{y}_j \tag{4.2.11}$$

Using $\boldsymbol{S}$ in place of $\boldsymbol{\Sigma}$ in (4.2.8) an estimated generalized least squares (EGLS) estimator of $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta}_e = [\boldsymbol{W}' (\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n) \boldsymbol{W}]^{-1} \boldsymbol{W}' (\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n) \boldsymbol{y} \tag{4.2.12}$$

This estimator does not have the same properties as $\boldsymbol{\theta}$. Unlike $\boldsymbol{\theta}$ in (4.2.8) $\boldsymbol{\theta}_e$ no longer has the minimum variance property. From (4.2.9), an approximate expression for the estimated variance-covariance matrix of $\boldsymbol{\theta}_e$ is given by

$$\hat{V}ar(\boldsymbol{\theta}_e) = [\boldsymbol{W}' (\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n) \boldsymbol{W}]^{-1} \tag{4.2.13}$$

Furthermore if $\boldsymbol{X}_i = \boldsymbol{X}_s$ and $\boldsymbol{Z}_i = \boldsymbol{Z}_s$, which means that $\boldsymbol{W}_i = \boldsymbol{W}_s$ for $i = 1, 2, \ldots, r$ then model (4.2.7) takes the form

$$\boldsymbol{y} = (\boldsymbol{I}_r \otimes \boldsymbol{W}_s) \boldsymbol{\theta} + \boldsymbol{\epsilon} \tag{4.2.14}$$

In this case, the GLS estimator of $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\theta} = \{\boldsymbol{I}_r \otimes [(\boldsymbol{W}_s' \boldsymbol{W}_s)^{-1} \boldsymbol{W}_s']\} \boldsymbol{y} \tag{4.2.15}$$

which is the form of the ordinary least squares (OLS) estimator of $\boldsymbol{\theta}$.

4.3 Estimating the Mean Response in the Presence of Fixed Block Effects

In Section 3.1 estimators of the model parameters was obtained by using the GLS method. In this section we derive the estimators of the mean responses by applying the same approach and by using the results in (4.2.8), (4.2.9), and (4.2.12).

Suppose that the ith response surface model is a polynomial of order d_i of the form

$$\eta_i(\boldsymbol{x}) = \beta_{0i} + \boldsymbol{f}_i'(\boldsymbol{x})\boldsymbol{\beta}_i \tag{4.3.1}$$

where $\boldsymbol{x} = (x_1\ x_2,\ \ldots\ ,x_k)'$ is a vector of k input variables $\eta_i(\boldsymbol{x})$ denotes the mean of the ith response at $\boldsymbol{x}$, and $\boldsymbol{f}_i'(\boldsymbol{x})$ is a vector whose elements consists of the x_i's defining the polynomial terms up to degree d_i. The elements of $\boldsymbol{\beta}_i$ and β_{0i} are unknown parameters. The least-squares estimator $\hat{\eta}_i(\boldsymbol{x})$ of the ith response at $\boldsymbol{x}$ is

$$\hat{\eta}_i(\boldsymbol{x}) = \hat{\beta}_{0i} + \boldsymbol{f}_i'(\boldsymbol{x})\hat{\boldsymbol{\beta}}_i \tag{4.3.2}$$

where $\hat{\beta}_{0i} = \frac{1}{b_i}\mathbf{1}_{b_i}'\boldsymbol{\tau}_i$. Based on this result, formula (4.3.2) can be written as

$$\hat{\eta}_i(\boldsymbol{x}) = \boldsymbol{g}_i'(\boldsymbol{x})\hat{\boldsymbol{\theta}}_i \tag{4.3.3}$$

where $\hat{\boldsymbol{\theta}}_i$ is the unbiased estimator of $\boldsymbol{\theta}_i$ obtained from (4.2.8) and $\boldsymbol{g}_i'(\boldsymbol{x}) = [\boldsymbol{f}_i'(\boldsymbol{x})\ \ \frac{1}{b_i}\mathbf{1}_{b_i}']$. Let $\hat{\boldsymbol{\eta}}(\boldsymbol{x}) = (\hat{\eta}_1(\boldsymbol{x})\ \ \hat{\eta}_2(\boldsymbol{x})\ \ \ldots\ \ \hat{\eta}_r(\boldsymbol{x}))'$. Using (4.2.9), the variance-covariance matrix of $\hat{\boldsymbol{\eta}}(\boldsymbol{x})$ is

$$Var[\hat{\boldsymbol{\eta}}(\boldsymbol{x})] = \boldsymbol{G}'[\boldsymbol{W}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{W}]^{-1}\boldsymbol{G} \tag{4.3.4}$$

where $\boldsymbol{G} = diag(\boldsymbol{g}_1(\boldsymbol{x}), \boldsymbol{g}_2(\boldsymbol{x})\ \ \ldots\ \ \boldsymbol{g}_r(\boldsymbol{x}))$. If $\boldsymbol{\Sigma}$ is unknown, we use the EGLS estimator $\hat{\boldsymbol{\theta}}_e$ given in equation (4.2.12) that is

$$\hat{\eta}_e(\boldsymbol{x}) = \boldsymbol{G}'\ \hat{\boldsymbol{\theta}}_e \tag{4.3.5}$$

and its estimated variance-covariance matrix is approximately given by

$$\hat{Var}[\hat{\boldsymbol{\eta}}_e(\boldsymbol{x})] = \boldsymbol{G}'[\boldsymbol{W}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{W}]^{-1}\boldsymbol{G} \tag{4.3.6}$$

4.4 Effect of Blocking on the Prediction Variance

Khuri (1994) pointed out in the single-response case that the block effects cause an increase in the prediction variance, assuming that the variance of the experimental error term in the model is the same whether or not block effects are included in the model, under the condition that block effects potentially exist. Specifically, if block effects exist it is assumed here that the variance of the experimental error term in the model with blocking terms is the same as the variance of the experimental error term in the model without blocking terms. Khuri (1994) showed that for a particular response, say response i,

$$Var[\hat{\eta}_i(\boldsymbol{x})] \geq Var[\hat{\eta}_{i0}(\boldsymbol{x})]$$

where $\eta_i(\boldsymbol{x})$ follows the form given in (4.3.3) and $\hat{\eta}_{i0}(\boldsymbol{x})$ is the least-squares estimator of the ith response at $\boldsymbol{x}$ when the block effects are zero. He proved this inequality by establishing the identity given by

$$Var[\hat{\eta}_i(\boldsymbol{x})] = Var[\hat{\eta}_{i0}(\boldsymbol{x})] + \boldsymbol{g}_i'(\boldsymbol{x})\boldsymbol{Q}^*\boldsymbol{g}_i(\boldsymbol{x})\sigma^2 \qquad (4.4.1)$$

In the above equation, $Var[\hat{\eta}_{i0}(\boldsymbol{x})]$ follows the form $c\sigma^2$, where $c = \boldsymbol{h}_i'(\boldsymbol{x})(\boldsymbol{U}_i'\boldsymbol{U}_i)^{-1}\boldsymbol{h}_i(\boldsymbol{x})$, $\boldsymbol{h}_i'(\boldsymbol{x}) = [1 \;\; \boldsymbol{f}_i'(\boldsymbol{x})]$, $\boldsymbol{U}_i = [\mathbf{1}_n \;\; \boldsymbol{X}_i]$, $\boldsymbol{g}_i'(\boldsymbol{x})$ is given in (4.3.3), $\sigma^2\boldsymbol{I}_n$ is the variance-covariance matrix of $\boldsymbol{\epsilon}_i$ in (4.2.4), $\boldsymbol{Q}^*$ is given by

$$\boldsymbol{Q}^* = (\boldsymbol{W}_i'\boldsymbol{W}_i)^{-1}\boldsymbol{K}_i[\boldsymbol{K}_i'(\boldsymbol{W}_i'\boldsymbol{W}_i)^{-1}\boldsymbol{K}_i]^{-1}\boldsymbol{K}_i'(\boldsymbol{W}_i'\boldsymbol{W}_i)^{-1}, \qquad (4.4.2)$$

$\boldsymbol{W}_i$ is defined in (4.2.5), $\boldsymbol{K}_i' = [\mathbf{0} \;\; \boldsymbol{L}_i]$, $\boldsymbol{L}_i = [\mathbf{1}_{b_i-1} \;\; -\boldsymbol{I}_{b_i-1}]$, and $\mathbf{0}$ is a zero matrix of order $(b_i - 1) \times p_i$, with p_i being the number of columns of $\boldsymbol{X}_i$. This implies that a model containing blocking terms will have a prediction variance value that is greater than or equal to the prediction variance for a model without blocking terms since $\boldsymbol{Q}^*$ is positive semidefinite. Thus the addition of the blocking terms will not reduce the prediction variance, assuming that the variance of the error term for the blocked data

is the same as the variance of the error term for the unblocked data In this section we demonstrate that the same is true for the multiresponse case

Let $\boldsymbol{\delta} = (\boldsymbol{\delta}_1' \ \boldsymbol{\delta}_2' \ \ldots \ \boldsymbol{\delta}_r')'$, where $\boldsymbol{\delta}_i$ is the vector of block effects for the ith response Let $Var[\boldsymbol{\eta}_0(\boldsymbol{x})]$ be the prediction variance when the block effects are zero that is, the elements of $\boldsymbol{\delta}$ are zero We will show that

$$Var[\boldsymbol{\eta}(\boldsymbol{x})] \geq Var[\boldsymbol{\eta}_0(\boldsymbol{x})]$$

for all $\boldsymbol{x}$ in a region of interest $\Re$ that is, $Var[\boldsymbol{\eta}(\boldsymbol{x})] - Var[\boldsymbol{\eta}_0(\boldsymbol{x})]$ is positive semidefinite for all $\boldsymbol{x} \in \Re$. To show this we note that if $\boldsymbol{\delta} = \boldsymbol{0}$ then $\boldsymbol{\tau}_i = \beta_{0i}\mathbf{1}_{k_i}$, $i = 1, 2, \ldots, r$ This means that the elements of $\boldsymbol{\tau}_i$ $(i = 1, 2, \ldots, r)$ are equal. We can therefore write $\boldsymbol{L}_i\boldsymbol{\tau}_i = \boldsymbol{0}$ where $\boldsymbol{L}_i = [\mathbf{1}_{k_i - 1} \ -\boldsymbol{I}_{k_i - 1}]$ Also the prediction equation for the ith response can be expressed as $\eta_{in}(\boldsymbol{x}) = \boldsymbol{g}_i'(\boldsymbol{x})\,\boldsymbol{\theta}_{oi}$ where $\boldsymbol{\theta}_{oi}$ is the least squares estimator of $\boldsymbol{\theta}_i$ under the restriction $\boldsymbol{L}_i\boldsymbol{\tau}_i = \boldsymbol{0}$ or equivalently $\boldsymbol{K}_i'\boldsymbol{\theta}_i = \boldsymbol{0}$ where $\boldsymbol{K}_i'$ is defined in (4.4.2) Hence $\boldsymbol{K}'\boldsymbol{\theta} = \boldsymbol{0}$ where $\boldsymbol{K}' = diag(\boldsymbol{K}_1' \ \boldsymbol{K}_2' \ \ldots \ \boldsymbol{K}_r')$ Now using a result given by Seber(1984 formula 8.28 p 403), the generalized least squares estimator of $\boldsymbol{\theta}$ under the above restriction is

$$\boldsymbol{\theta}_0 = \boldsymbol{\theta} - \boldsymbol{A}^{-1}\boldsymbol{K}(\boldsymbol{K}'\boldsymbol{A}^{-1}\boldsymbol{K})^{-1}\boldsymbol{K}'\boldsymbol{\theta} \tag{4.4.3}$$

where $\boldsymbol{A} = \boldsymbol{W}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{W}$ From this result the variance-covariance matrix of $\boldsymbol{\theta}_0$ is

$$Var(\boldsymbol{\theta}_0) = \boldsymbol{P}\boldsymbol{A}^{-1}\boldsymbol{P}' \tag{4.4.4}$$

where

$$\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{A}^{-1}\boldsymbol{K}(\boldsymbol{K}'\boldsymbol{A}^{-1}\boldsymbol{K})^{-1}\boldsymbol{K}'$$

This variance-covariance matrix can also be written in the form

$$\begin{aligned} Var(\boldsymbol{\theta}_0) \ = \ & \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{K}(\boldsymbol{K}'\boldsymbol{A}^{-1}\boldsymbol{K})^{-1}\boldsymbol{K}'\boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{K}(\boldsymbol{K}'\boldsymbol{A}^{-1}\boldsymbol{K})^{-1}\boldsymbol{K}'\boldsymbol{A}^{-1} \\ & + \boldsymbol{A}^{-1}\boldsymbol{K}(\boldsymbol{K}'\boldsymbol{A}^{-1}\boldsymbol{K})^{-1}\boldsymbol{K}'\boldsymbol{A}^{-1}\boldsymbol{K}(\boldsymbol{K}'\boldsymbol{A}^{-1}\boldsymbol{K})^{-1}\boldsymbol{K}'\boldsymbol{A}^{-1} \end{aligned}$$

$$
\begin{aligned}
&= A^{-1} - A^{-1}K(KA^{-1}K)^{-1}K'A^{-1} - A^{-1}K(K'A^{-1}K)^{-1}K'A^{-1} \\
&\quad + A^{-1}K(KA^{-1}K)^{-1}K'A^{-1} \\
&= A^{-1} - A^{-1}K(K'A^{-1}K)^{-1}K'A^{-1}
\end{aligned}
$$

The variance-covariance matrix of $\eta(x)$ in (4.3.4) can then be expressed as

$$
\begin{aligned}
Var[\eta(x)] &= G'A^{-1}G \\
&= G'[Var(\theta_0) + Q]G \\
&= Var[\eta_0(x)] + G'QG
\end{aligned}
$$

where $Q = A^{-1}K(K'A^{-1}K)^{-1}K'A^{-1}$ is of order $(p + \sum_{i=1}^{r} b_i) \times (p + \sum_{i=1}^{r} b_i)$, $p = \sum_{i=1}^{r} p_i$ and the rank of Q is equal to the rank of K, which is equal to $\sum_{i=1}^{r} b_i - r$. Since Q is positive semidefinite, it follows that $Var[\eta(x)] \geq Var[\eta_0(x)]$, assuming that Σ remains unchanged. However if Σ is replaced by its estimator S, then this inequality may not necessarily hold.

4.5 Orthogonal Blocking

It is often the case that the polynomial effects in the models described in equation (4.2.4) are not independent of the block effects. Box and Hunter (1957) pointed out that the manner in which the experimental units of a design are assigned to the blocks greatly affects the efficiency of the estimation of the model parameters. As a result, it would be desirable to analyze the polynomial effects independently of the block effects. We can achieve this by *orthogonal blocking*. A design in which the polynomial effects can be estimated independently of the block effects is said to block orthogonally. Thus finding conditions for orthogonal blocking for the multiresponse case is desired. Consider the models given in equation (4.2.4). The multiresponse form of these models is

$$
y = M\beta_0 + X\beta + Z\delta + \epsilon \tag{4.5.1}
$$

where $\boldsymbol{M} = \boldsymbol{I}_r \otimes \mathbf{1}_n$ $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02}, \quad \beta_{0r})'$, $\boldsymbol{X} = diag(\boldsymbol{X}_1\ \boldsymbol{X}_2 \quad \boldsymbol{X}_r)$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1'\ \boldsymbol{\beta}_2' \quad \boldsymbol{\beta}_r')'$, $\boldsymbol{Z} = diag(\boldsymbol{Z}_1\ \boldsymbol{Z}_2 \quad \boldsymbol{Z}_r)$ and $\boldsymbol{\delta} = (\boldsymbol{\delta}_1', \boldsymbol{\delta}_2', \quad \boldsymbol{\delta}_r')'$ In this case the polynomials used for the different responses can have different degrees d_i Furthermore the models given in equation (4 2 4) can also be written in the form

$$\boldsymbol{y}_i = \lambda_{0i}\mathbf{1}_n + \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{Z}_i\boldsymbol{\delta}_i + \boldsymbol{\epsilon}_i, \quad i = 1\ 2 \quad r \tag{4 5 2}$$

where $\lambda_{0i} = \beta_{0i} + \frac{1}{n}\mathbf{1}_n'\boldsymbol{Z}_i\boldsymbol{\delta}_i$ and $\boldsymbol{Z}_i = (\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)\boldsymbol{Z}_i$ The multiresponse form of the models described in equation (4 5 2) is

$$\boldsymbol{y} = \boldsymbol{M}\boldsymbol{\lambda}_0 + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\delta} + \boldsymbol{\epsilon} \tag{4 5 3}$$

where $\boldsymbol{\lambda}_0 = (\lambda_{01}, \lambda_{02}, \quad \lambda_{0r})'$, $\bar{\boldsymbol{Z}} = diag(\boldsymbol{Z}_1, \bar{\boldsymbol{Z}}_2 \quad \boldsymbol{Z}_r)$, and $\boldsymbol{M}$ $\boldsymbol{X}$, $\boldsymbol{\delta}$, and $\boldsymbol{\epsilon}$ are the same as in model (4 5.1) Model (4 5 3) can be rewritten as

$$\boldsymbol{y} = \boldsymbol{X}^0\boldsymbol{\beta}^0 + \boldsymbol{Z}\boldsymbol{\delta} + \boldsymbol{\epsilon} \tag{4 5 4}$$

where $\boldsymbol{X}^0 = (\boldsymbol{M} \quad \boldsymbol{X})$ and $\boldsymbol{\beta}^0 = (\boldsymbol{\lambda}_0 \quad \boldsymbol{\beta}')'$

For orthogonal blocking the generalized least squares estimator of $\boldsymbol{\beta}^0$ must be independent of $\boldsymbol{\delta}$ On this basis, a design associated with model (4 5.4) blocks orthogonally if and only if

$$\boldsymbol{X}^{0\prime}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\bar{\boldsymbol{Z}} = \boldsymbol{O} \tag{4 5 5}$$

Now let $\boldsymbol{\Sigma}^{-1} = (\sigma^{ij})$ $i\ j = 1, 2, \quad r$ Then equation (4 5 5) can be expressed as

$$\begin{pmatrix} \sigma^{11}\mathbf{1}_n'\boldsymbol{Z}_1 & \sigma^{12}\mathbf{1}_n'\boldsymbol{Z}_2 & \sigma^{1r}\mathbf{1}_n'\boldsymbol{Z}_r \\ \sigma^{21}\mathbf{1}_n'\boldsymbol{Z}_1 & \sigma^{22}\mathbf{1}_n'\boldsymbol{Z}_2 & \sigma^{2r}\mathbf{1}_n'\boldsymbol{Z}_r \\ \\ \sigma^{r1}\mathbf{1}_n'\boldsymbol{Z}_1 & \sigma^{r2}\mathbf{1}_n'\boldsymbol{Z}_2 & \sigma^{rr}\mathbf{1}_n'\boldsymbol{Z}_r \\ \sigma^{11}\boldsymbol{X}_1'\boldsymbol{Z}_1 & \sigma^{12}\boldsymbol{X}_1'\boldsymbol{Z}_2 & \sigma^{1r}\boldsymbol{X}_1'\boldsymbol{Z}_r \\ \sigma^{21}\boldsymbol{X}_2'\boldsymbol{Z}_1 & \sigma^{22}\boldsymbol{X}_2'\boldsymbol{Z}_2 & \sigma^{2r}\boldsymbol{X}_2\boldsymbol{Z}_r \\ \\ \sigma^{r1}\boldsymbol{X}_r'\boldsymbol{Z}_1 & \sigma^{r2}\boldsymbol{X}_r'\boldsymbol{Z}_2 & \sigma^{rr}\boldsymbol{X}_r'\boldsymbol{Z}_r \end{pmatrix} = \boldsymbol{O} \tag{4 5 6}$$

Thus if condition (4 5 5) or (4 5 6) is satisfied it is possible to estimate the polynomial effects in multiresponse model (4 5 4) independently of the block effects. Based on this condition for orthogonal blocking, two theorems one on the estimation of model parameters and the other on the prediction variance will be established. The following results are needed to prove these theorems. From equation (4 5 6), the matrix

$$\begin{pmatrix} \sigma^{11}\mathbf{1}_n'(\boldsymbol{J}_n/n)\boldsymbol{Z}_1 & \sigma^{12}\mathbf{1}_n'(\boldsymbol{J}_n/n)\boldsymbol{Z}_2 & \sigma^{1r}\mathbf{1}_n'(\boldsymbol{J}_n/n)\boldsymbol{Z}_r \\ \sigma^{21}\mathbf{1}_n'(\boldsymbol{J}_n/n)\boldsymbol{Z}_1 & \sigma^{22}\mathbf{1}_n'(\boldsymbol{J}_n/n)\boldsymbol{Z}_2 & \sigma^{2r}\mathbf{1}_n'(\boldsymbol{J}_n/n)\boldsymbol{Z}_r \\ \\ \sigma^{r1}\mathbf{1}_n'(\boldsymbol{J}_n/n)\boldsymbol{Z}_1 & \sigma^{r2}\mathbf{1}_n'(\boldsymbol{J}_n/n)\boldsymbol{Z}_2 & \sigma^{rr}\mathbf{1}_n'(\boldsymbol{J}_n/n)\boldsymbol{Z}_r \\ \sigma^{11}\boldsymbol{X}_1'(\boldsymbol{J}_n/n)\boldsymbol{Z}_1 & \sigma^{12}\boldsymbol{X}_1'(\boldsymbol{J}_n/n)\boldsymbol{Z}_2 & \sigma^{1r}\boldsymbol{X}_1'(\boldsymbol{J}_n/n)\boldsymbol{Z}_r \\ \sigma^{21}\boldsymbol{X}_2'(\boldsymbol{J}_n/n)\boldsymbol{Z}_1 & \sigma^{22}\boldsymbol{X}_2'(\boldsymbol{J}_n/n)\boldsymbol{Z}_2 & \sigma^{2r}\boldsymbol{X}_2'(\boldsymbol{J}_n/n)\boldsymbol{Z}_r \\ \\ \sigma^{r1}\boldsymbol{X}_r'(\boldsymbol{J}_n/n)\boldsymbol{Z}_1 & \sigma^{r2}\boldsymbol{X}_r'(\boldsymbol{J}_n/n)\boldsymbol{Z}_2 & \sigma^{rr}\boldsymbol{X}_r'(\boldsymbol{J}_n/n)\boldsymbol{Z}_r \end{pmatrix} \tag{4 5 7}$$

is equal to

$$\begin{pmatrix} \boldsymbol{M}'(\Sigma^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z} \\ \boldsymbol{X}'(\Sigma^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z} \end{pmatrix} \tag{4 5 8}$$

Let $\boldsymbol{\mathcal{J}} = \boldsymbol{I}_r \otimes \boldsymbol{J}_n$. From (4.5 7) and (4.5 8) we have

$$\boldsymbol{X}'(\Sigma^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z} = \boldsymbol{X}'(\Sigma^{-1} \otimes \boldsymbol{I}_n)\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z} \tag{4.5 9}$$

If condition (4 5 5) is satisfied, some important results can be obtained. To be able to derive these results, we first consider the models described in equation (4 2.5), that is,

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{Z}_i\boldsymbol{\tau}_i + \boldsymbol{\epsilon}_i, \quad i = 1, 2, \ldots, r$$

where $\boldsymbol{\tau}_i = \beta_{0i}\mathbf{1}_b + \boldsymbol{\delta}_i$. The multiresponse form of the above models is then

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\tau} + \boldsymbol{\epsilon} \tag{4 5.10}$$

where $\boldsymbol{\tau} = (\boldsymbol{\tau}_1', \boldsymbol{\tau}_2', \ldots, \boldsymbol{\tau}_r')'$ and $\boldsymbol{X}$, $\boldsymbol{\beta}$, $\boldsymbol{Z}$ and $\boldsymbol{\epsilon}$ are as defined in model (4 5 1). Note that the latter model is the same as

$$\boldsymbol{y} = \boldsymbol{X}^*\boldsymbol{\theta} + \boldsymbol{Z}\boldsymbol{\delta} + \boldsymbol{\epsilon} \tag{4 5 11}$$

where $\boldsymbol{X}^{e}$ is as defined in equation (4 5 4) and $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\beta}}_0' \quad \tilde{\boldsymbol{\beta}}')'$

As stated earlier orthogonal blocking allows us to estimate the polynomial effects independently of the block effects. In the following theorem, we also show that orthogonal blocking allows us to get correct estimates for the polynomial effects even if the block effects were incorrectly excluded in the analysis of the multiresponse experiment

Theorem 4 5 1 If the design blocks orthogonally then the generalized least squares estimator $\tilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ in model (4 5 10) is the same as when the block effects $\boldsymbol{\delta}$ are zero that is $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$ where $\hat{\boldsymbol{\beta}}$ satisfies

$$\tilde{\boldsymbol{\theta}} = \begin{pmatrix} \tilde{\boldsymbol{\beta}}_0 \\ \tilde{\boldsymbol{\beta}} \end{pmatrix} = [\boldsymbol{X}^{e\prime}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}^{e}]^{-1}\boldsymbol{X}^{e\prime}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} \qquad (4\,5\,12)$$

Proof The proof is given in Appendix D Section D 1

The importance of the above theorem lies in the fact that even if an experimenter incorrectly analyzes a multiresponse experiment with block effects by excluding them in the analysis the resulting GLS estimator for $\boldsymbol{\beta}$ would still be correct provided that the design used blocks orthogonally The other important result for orthogonal blocking is established in Theorem 4 5 2. In Section 4 it was pointed out that the block effects cause an increase in the prediction variance. In the following theorem we show that when the design blocks orthogonally the prediction variance at a point $\boldsymbol{x}$ inside the experimental region exceeds $Var[\boldsymbol{\eta}_0(\boldsymbol{x})]$ by a constant amount which depends on the sizes of the blocks for a given $\boldsymbol{\Sigma}$

Theorem 4.5.2 If the design blocks orthogonally and if the blocks are the same for all the r responses of interest, that is $\boldsymbol{Z}_1 = \boldsymbol{Z}_2 = \cdots = \boldsymbol{Z}_r = \boldsymbol{Z}_0$, then the prediction variance given in equation (4 3 4) takes the form

$$Var[\boldsymbol{\eta}(\boldsymbol{x})] = Var[\boldsymbol{\eta}_0(\boldsymbol{x})] + \left[\frac{1}{b^2}\left(\sum_{i=1}^{b}\frac{1}{n_i}\right) - \frac{k}{n}\right]\boldsymbol{\Sigma} \qquad (4\,5\,13)$$

where $Var[\boldsymbol{\eta}_0(\boldsymbol{x})]$ *denotes the prediction variance when the block effects* $\boldsymbol{\delta} = \mathbf{0}$, *that is*

$$Var[\boldsymbol{\eta}_0(\boldsymbol{x})] = \boldsymbol{H}\,[\boldsymbol{X}^{o\prime}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}^{o}]^{-1}\boldsymbol{H} \tag{4.5.14}$$

where $\boldsymbol{H}' = [\boldsymbol{I}_r \quad \boldsymbol{F}^{o}]$, $\boldsymbol{F}^{o} = diag(\boldsymbol{f}_1'(\boldsymbol{x}) \;\; \boldsymbol{f}_2'(\boldsymbol{x}) \;\; \ldots \;\; \boldsymbol{f}_r'(\boldsymbol{x}))$ *and* $\boldsymbol{f}_i'(\boldsymbol{x})$ *is as defined in model (4.3.1).*

Proof: The proof is given in Appendix D, Section D.2.

Note that to apply Theorem 4.5.2, we need not assume that the polynomial models for the responses have the same degree, that is, the polynomial models can have different degrees d_i.

4.6 A Numerical Example

In this example, we use the results of an experiment performed by Chen et al. (1989) concerning the post-storage condition of freshly harvested Delicious apples. The apples were harvested from the same orchards at two different dates, that is, Harvest I: is on day 1 of the commercial harvesting date and Harvest II: is 14 days after. We consider these two harvesting dates as the fixed block effects. In the experiment, The input variables considered were X_1 = number of weeks in storage after harvest and X_2 = storage temperature (^oC) after harvest. There were three response variables of interest: y_1 = flesh firmness (N), y_2 = amount of extractable juice (ml/100g fluid weight), and y_3 = internal ethylene composition of the apples (ppm). A 4×4 factorial design was used. The design settings in the original and coded variables and the multiresponse data are given in Table 4.1. Note that the square root of the y_3 values are given in Table 4.1 and that the coded values of X_1 and X_2 are derived from the following equations.

$$x_1 = \frac{X_1 - 2.5}{1.12}$$

Table 4 1 Experimental Design and Response Values

	Original Input Variables		Coded Input Variables		Responses		
HARVEST	X_1	X_2	x_1	x_2	y_1	y_2	$(y_1)^2$
I	1	0 0	-1 34	-1 21	62 41	75	2 2694
I	2	0 0	0 45	-1 21	63 16	73	2 4269
I	3	0 0	0 45	-1 21	60 34	75	2 7129
I	4	0 0	1 34	-1 21	60 80	74	3 5114
I	1	4 4	-1 34	-0 64	57 74	75	4 2024
I	2	4 4	0 45	-0 64	58 18	73 5	4 1593
I	3	4 4	0 45	-0 64	58 89	72	4 8908
I	4	4 4	1.34	0 64	55 29	73	3 5369
I	1	12.8	-1 34	0 45	57 71	73	7 1777
I	2	12 8	-0 45	0 45	51 77	72	8 1841
I	3	12 8	0 45	0 45	41 77	71	11 7631
I	4	12.8	1 34	0 45	38 25	70 5	10 0050
I	1	20 0	-1 34	1 38	53 62	74	10 1509
I	2	20 0	-0 45	1 38	43 59	71	11 1200
I	3	20 0	0 45	1 38	36 52	70	12 0108
I	4	20.0	1 34	1 38	32.00	68	11 7004
II	1	0 0	-1 34	-1 21	61 38	74 5	1 8708
II	2	0 0	0 45	1 21	59 65	75	3 0033
II	3	0 0	0 45	1 21	58 34	73	2 6439
II	4	0 0	1 34	-1 21	57 02	74	2 6796
II	1	4 4	1 34	0 64	60 85	74	3.0935
II	2	4 4	0 45	0 64	56 22	73	5 4936
II	3	4 4	0 45	-0 64	57 68	72	5 2536
II	4	4 4	1 34	-0 64	57 02	73	4 9285
II	1	12 8	-1 34	0 45	52 06	72	7.3797
II	2	12 8	0 45	0 45	47 46	71	11 5100
II	3	12 8	0 45	0 45	39.94	70	10 6461
II	4	12 8	1 34	0 45	36 45	69	12 0677
II	1	20 0	1 34	1 38	52 22	72	7 7685
II	2	20 0	0 45	1 38	44 75	72 5	12 4912
II	3	20 0	0 45	1 38	41 54	69	10 0050
II	4	20 0	1 34	1 38	32.92	68	12 0108

$$\frac{\bar{Y}_2 - 9.5}{7.70}$$

2.5 and 9.3 are the means of Y_1 and Y_2, respectively, and 1.12 and 7.70 are corresponding standard deviations where the formula $\sqrt{\frac{\sum_{j=1}^{n}(X_j - \bar{X}_i)^2}{n}}$ $i = 1, 2,$ was used. From Table 4.1 we can see that all the responses have the same blocking arrangement.

For this example, we wish to determine optimum operating conditions on storage time and temperature for both harvest times. To do so, we first need to determine the best multiresponse model, taking into account the fixed block effects. Once the model is established, we can use the optimization procedure given in Chapter 3 to obtain the simultaneous optima of the three responses.

Using the C_p, R^2, and adjusted R^2 criteria described in Chapter 3, the "best" model for each response was determined. The procedure PROC RSQUARE of SAS was used to determine these models and the results are given in Appendix E. A square root transformation was used to stabilize the residuals for y_3. This means that the vector $\boldsymbol{y}_3$ used in the model and parameter estimation contains the square roots of the observed values for y_3. A complete second-order model in x_1 and x_2 was considered to be the best for each of the three responses based on the C_p, R^2, and adjusted R^2 values given in Appendix D. Applying equation (4.2.5), the "best" model for all three responses can be written in the form

$$\boldsymbol{y}_i = \boldsymbol{W}_e\boldsymbol{\theta}_i + \boldsymbol{\epsilon}_i, \quad i = 1, 2, 3$$

where $\boldsymbol{W}_e = [\boldsymbol{X}_e \quad \boldsymbol{Z}_e]$, $\boldsymbol{Z}_e = diag(\mathbf{1}_{16}, \mathbf{1}_{16})$, and $\boldsymbol{X}_e$ is a matrix whose uth row is given by $\boldsymbol{f}'(\boldsymbol{x}_u) = (x_{1u}, x_{2u}, x_{1u}^2, x_{2u}^2, x_{1u}x_{2u})$, $u = 1, 2, \ldots, 32$. Using equation (4.2.14), the multiresponse form of these models is given by

$$\boldsymbol{y} = (\boldsymbol{I}_r \otimes \boldsymbol{W}_e)\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

Table 4.1 Cp, R^2 and adjusted R^2 values

Criterion	η_1	η_2	η_3
R^2	0.936	0.979	0.930
adjusted R^2	0.923	0.975	0.916
Cp	6	6	6

where $\boldsymbol{y} = (\boldsymbol{y}_1' \ \boldsymbol{y}_2' \ \boldsymbol{y}_3')'$ $\boldsymbol{\theta} = (\boldsymbol{\theta}_1' \ \boldsymbol{\theta}_2' \ \boldsymbol{\theta}_3')'$, and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1', \boldsymbol{\epsilon}_2', \boldsymbol{\epsilon}_3')$. Note that since $\boldsymbol{W}_1 = \boldsymbol{W}_2 = \boldsymbol{W}_3 = \boldsymbol{W}_e$, the GLS estimator of $\boldsymbol{\theta}$ does not depend on $\boldsymbol{\Sigma}$ that is, the GLS estimator coincides with the OLS estimator. Srivastava and Giles (1987, p.17, equations 2.15-2.16) showed that in this case, the EGLS estimator is the same as the OLS estimator. Thus using the OLS estimator, $\hat{\boldsymbol{\theta}}_i = (\boldsymbol{W}_e'\boldsymbol{W}_e)^{-1}\boldsymbol{W}_e'\boldsymbol{y}_i$, for $i = 1, 2, 3$ we have

$$\hat{\boldsymbol{\theta}}_1 = (-4.13, -7.79, 0.31, 1.05, -2.85, 50.61, 49.57)'$$

$$\hat{\boldsymbol{\theta}}_2 = (-0.96, -1.53, 0.24, 0.42, -0.61, 71.84, 71.34)'$$

$$\hat{\boldsymbol{\theta}}_3 = (0.72, 3.57, -0.54, -0.87, 0.33, 8.30, 8.49)'$$

Using equation (4.3.3), the least squares estimates of $\eta_i(\boldsymbol{x})$ are

$$\hat{\eta}_1(\boldsymbol{x}) = 50.09 - 4.13x_1 - 7.79x_2 + 0.31x_1^2 + 1.05x_2^2 - 2.85x_1x_2$$

$$\hat{\eta}_2(\boldsymbol{x}) = 71.59 - 0.96x_1 - 1.53x_2 + 0.24x_1^2 + 0.42x_2^2 - 0.61x_1x_2$$

$$\hat{\eta}_3(\boldsymbol{x}) = 8.395 + 0.72x_1 + 3.57x_2 - 0.54x_1^2 - 0.87x_2^2 + 0.33x_1x_2$$

Note that $\hat{\eta}_3(\boldsymbol{x})$ here corresponds to the square root transformation used for y_3. The Cp, R^2 and adjusted R^2 for the above models are given in Table 4.2.

It is desired to find the operating conditions that maximize y_1 and y_2 and minimize y_3. For this purpose we can apply the simultaneous optimization procedure given in Chapter 3. This amounts to minimizing the distance function given by

$$\rho[\boldsymbol{\eta}(\boldsymbol{x}), \boldsymbol{\nu}] = \left[\frac{[\boldsymbol{\eta}(\boldsymbol{x}) - \boldsymbol{\nu}]'\boldsymbol{\Sigma}^{-1}[\boldsymbol{\eta}(\boldsymbol{x}) - \boldsymbol{\nu}]}{\boldsymbol{g}'(\boldsymbol{x})(\boldsymbol{W}_e'\boldsymbol{W}_e)^{-1}\boldsymbol{g}(\boldsymbol{x})}\right]^{\frac{1}{2}}$$

where $\boldsymbol{\nu}$ is the $r \times 1$ estimated vector of individual optima over the experimental region $\mathcal{R}$ given by the inequalities $-1.34 \leq x_1 \leq 1.34$ and $-1.21 \leq x_2 \leq 1.38$, $\boldsymbol{g}'(\boldsymbol{x})$ is a vector of order $1 \times p_o$ that has the same form as a row of $\boldsymbol{W}_o$ but is evaluated at the point $\boldsymbol{x}$, p_o is the number of columns of $\boldsymbol{W}_o$, and $\boldsymbol{\Sigma} = \frac{1}{n-p_o}\boldsymbol{Y}'[\boldsymbol{I}_n - \boldsymbol{W}_o(\boldsymbol{W}_o'\boldsymbol{W}_o)^{-1}\boldsymbol{W}_o']\boldsymbol{Y}$ where $\boldsymbol{Y}$ is the multiresponse data matrix

$$\boldsymbol{Y} = [\boldsymbol{y}_1 \quad \boldsymbol{y}_2 \quad \boldsymbol{y}_3],$$

and n is the total number of observations. Note that in this example $n = 32$ and $p_o = 7$. Using a FORTRAN program written by Conlon (1992) (see Appendix 2.3), this distance function is minimized and the simultaneous optima of all three responses over $\mathcal{R}$ are

$$\boldsymbol{\eta}_{opt} = \begin{pmatrix} 62.53 \\ 74.79 \\ 1.39 \end{pmatrix}$$

The minimum of η_3 given above is the one obtained by using the square root transformation to find the minimum in terms of the original scale of measurement of y_3; we simply square 1.39 and get 1.93. These simultaneous optima are attained at the location $x_{1opt} = -1.34$ and $x_{2opt} = -1.21$, that is a holding period of 1 week at 0°C. At this location we have

$$\begin{aligned} Var[\hat{\eta}_1(\boldsymbol{x}_{opt})] &= 1.47, \\ Var[\hat{\eta}_2(\boldsymbol{x}_{opt})] &= 0.11 \\ Var[\hat{\eta}_3(\boldsymbol{x}_{opt})] &= 0.30 \end{aligned}$$

The design used in this example blocks orthogonally. From Theorem 2.5.2 we then have

$$Var[\boldsymbol{\eta}(\boldsymbol{x})] = Var[\boldsymbol{\eta}_o(\boldsymbol{x})] + \left[\frac{1}{2^2}\sum_{i=1}^{2}\frac{1}{16} - \frac{1}{32}\right]\boldsymbol{\Sigma}$$

The second term in the above equation is equal to a zero matrix since $\frac{1}{n^2}\sum_{i=1}^{b}\frac{1}{1b}-\frac{1}{n^2}=0$ This means that there is no increase in the prediction variance due to blocking In general, using Theorem 4 5 2 we can say that

$$\mathbb{V}ar[\boldsymbol{\eta}(\boldsymbol{x})]=\mathbb{V}ar[\boldsymbol{\eta}_a(\boldsymbol{x})]$$

provided that $n_1=n_2=\quad=n_b$ To prove this we note that if $n_1=n_2=\quad=n_b=n^*$ then the quantity given by $\left[\frac{1}{b^2}\left(\sum_{i=1}^{b}\frac{1}{n_i}\right)-\frac{1}{n}\right]$ simplifies to 0

CHAPTER 5

MULTIRESPONSE MODELS WITH RANDOM BLOCK EFFECTS

5.1 Introduction

In Chapter 4 a discussion of multiresponse models with block effects was given with the assumption that the block effects being considered are fixed. It is not always the case that the block effects in an experiment are fixed. There are many experimental situations in which the block effects should be treated as random. For example, in a chemical process the raw material used may come in batches in which the quality of the material vary from batch to batch. In this case, when we consider this extraneous source of variation, the block effect (batches) should be treated as random. In this chapter we consider the analysis of multiresponse models with random block effects. It is important to correctly identify the nature of the block effects being considered because conducting a fixed-effects analysis when the block effect is random may lead to incorrect conclusions. This was shown by Khuri (1992) in the single-response case. We shall also consider interactions between the fixed polynomial effects and the random block effects, that is, the block × treatment interactions, in the analysis. The approach developed in Chapter 4 is used to determine the multiresponse models and to obtain estimators of the model parameters. The results given by Khuri (1992, 1994, 1996b) are applied and extended to the multiresponse case.

In this section we follow the approach described in Chapter 4, Section 4.2 to establish the multiresponse model and to determine estimators for the model parameters. Consider the same models as in Section 3.2, namely

$$\boldsymbol{y}_i = \beta_{0i}\mathbf{1}_n + \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{Z}_i\boldsymbol{\delta}_i + \boldsymbol{\epsilon}_i \quad i = 1, 2, \ldots, r \tag{5.2.1}$$

Assume that all r responses have the same number of observations, n, and that b blocks are used for each response. Suppose that the block effects are random such that each $\boldsymbol{\delta}_i$, $i = 1, 2, \ldots, r$ has a normal distribution with mean $E(\boldsymbol{\delta}_i) = \mathbf{0}$ and $E(\boldsymbol{\delta}_i\boldsymbol{\delta}_j') = \sigma_{ij}\boldsymbol{I}_b$, $i \neq j$, $i, j = 1, 2, \ldots, r$, that is, for the random effects $\boldsymbol{\delta}_i = (\delta_{i1}, \delta_{i2}, \ldots, \delta_{ib})'$ and $\boldsymbol{\delta}_j = (\delta_{j1}, \delta_{j2}, \ldots, \delta_{jb})'$ where $i \neq j$ we have $E(\delta_{iu}\delta_{jv}) = \sigma_{ij}$, $u = 1, 2, \ldots, b$ and $E(\delta_{iu}\delta_{jv}) = 0$ $u \neq v$, $u, v = 1, 2, \ldots, b$. Furthermore the block effects are assumed to be independent of $\boldsymbol{\epsilon}_i$ which also has the normal distribution with mean $E(\boldsymbol{\epsilon}_i) = 0$ and $E(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_j') = \sigma_{ij}\boldsymbol{I}_n$, $i \neq j$, $i, j = 1, 2, \ldots, r$. Note that the above models can be rewritten as

$$\boldsymbol{y}_i = \beta_{0i}\mathbf{1}_n + \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{\nu}_i, \quad i = 1, 2, \ldots, r, \tag{5.2.2}$$

where $\boldsymbol{\nu}_i = \boldsymbol{Z}_i\boldsymbol{\delta}_i + \boldsymbol{\epsilon}_i$. Under the assumptions given above we have

$$E(\boldsymbol{\nu}_i) = \mathbf{0}, \tag{5.2.3}$$

and

$$E(\boldsymbol{\nu}_i\boldsymbol{\nu}_j') = \sigma_{ij}\boldsymbol{I}_n + \sigma_{ij}\boldsymbol{Z}_i\boldsymbol{Z}_j' = \boldsymbol{\Phi}_{ij} \tag{5.2.4}$$

The multiresponse form of the models given in equation (5.2.2) is

$$\boldsymbol{y} = \boldsymbol{M}\boldsymbol{\beta}_0 + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\nu} \tag{5.2.5}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1', \boldsymbol{y}_2', \ldots, \boldsymbol{y}_r')'$, $\boldsymbol{M} = \boldsymbol{I}_r \otimes \mathbf{1}_n$, $\boldsymbol{X} = diag(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_r)$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_r')'$, and $\boldsymbol{\nu} = (\boldsymbol{\nu}_1', \boldsymbol{\nu}_2', \ldots, \boldsymbol{\nu}_r')$. Equation (5.2.5) in turn can be rewritten

$$\boldsymbol{y} = \boldsymbol{X}^0\boldsymbol{\vartheta} + \boldsymbol{\nu} \qquad (5\ 2\ 6)$$

where $\boldsymbol{X}^0 = (\boldsymbol{M} \quad \boldsymbol{X})$ and $\boldsymbol{\vartheta} = (\beta_0' \ \boldsymbol{\beta}')'$ Let $\boldsymbol{\Sigma} = (\sigma_{ij})$ and $\boldsymbol{\Sigma}_\delta = (\sigma_{\delta ij})$, then from equations (5 2.3) and (5 2 4) we have

$$E(\boldsymbol{\nu}) = 0 \qquad (5\ 2\ 7)$$

and

$$E(\boldsymbol{\nu}\boldsymbol{\nu}') = (\boldsymbol{\Sigma} \otimes \boldsymbol{I}_n) + \boldsymbol{Z}(\boldsymbol{\Sigma}_\delta \otimes \boldsymbol{I}_h)\boldsymbol{Z}' = \boldsymbol{\Phi} \qquad (5\ 2\ 8)$$

where $\boldsymbol{Z} = diag(\boldsymbol{Z}_1 \ \boldsymbol{Z}_2, \quad \boldsymbol{Z}_\tau)$

To obtain estimates of the parameters in model (5 2-5) we use the generalized least squares (GLS) approach as described by Srivastava and Giles(1987, p 12) The GLS estimates are given by

$$\hat{\boldsymbol{\vartheta}} = (\boldsymbol{X}^{0\prime}\boldsymbol{\Phi}^{-1}\boldsymbol{X}^0)^{-1}\boldsymbol{X}^{0\prime}\boldsymbol{\Phi}^{-1}\boldsymbol{y} \qquad (5\ 2\ 9)$$

that is,

$$\hat{\boldsymbol{\vartheta}} = \begin{pmatrix} \beta_0 \\ \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \boldsymbol{M}'\boldsymbol{\Phi}^{-1}\boldsymbol{M} & \boldsymbol{M}'\boldsymbol{\Phi}^{-1}\boldsymbol{X} \\ \boldsymbol{X}'\boldsymbol{\Phi}^{-1}\boldsymbol{M} & \boldsymbol{X}'\boldsymbol{\Phi}^{-1}\boldsymbol{X} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{M}'\boldsymbol{\Phi}^{-1}\boldsymbol{y} \\ \boldsymbol{X}'\boldsymbol{\Phi}^{-1}\boldsymbol{y} \end{pmatrix} \qquad (5\ 2\ 10)$$

The parameter $\boldsymbol{\beta}$ is of interest therefore we will concentrate mainly on its estimate The GLS estimate of $\boldsymbol{\beta}$ from equation (5 2 10) can be obtained by using the result given by Seber(1984 p 519 formula A3 1) for finding the inverse of a partitioned matrix. Applying this result the GLS estimator is given by

$$\begin{aligned} \hat{\boldsymbol{\beta}} \ = \ & [\boldsymbol{X}'\boldsymbol{\Phi}^{-1}\boldsymbol{X} - \boldsymbol{X}'\boldsymbol{\Phi}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{\Phi}^{-1}\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{\Phi}^{-1}\boldsymbol{X}]^{-1}[\boldsymbol{X}'\boldsymbol{\Phi}^{-1}\boldsymbol{y} - \\ & \boldsymbol{X}'\boldsymbol{\Phi}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{\Phi}^{-1}\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{\Phi}^{-1}\boldsymbol{y}] \end{aligned} \qquad (5\ 2\ 11)$$

5.3 Effect of Orthogonal Blocking on the Estimation of Parameters

The condition for orthogonal blocking in the multiresponse case and its advantages were given in Section 4.5. This condition will be used in the present section to determine its effect on the GLS estimate of $\boldsymbol{\beta}$. The effect of orthogonal blocking on the parameter estimates was discussed by Khuri(1992) in the single-response case. He showed that if the design blocks orthogonally, then the GLS estimate of $\boldsymbol{\beta}$ when the block effects are random is equal to the GLS estimate of $\boldsymbol{\beta}$ when the block effects are fixed and both are equal to the OLS estimate of $\boldsymbol{\beta}$ (which ignores the block effect). We now extend the development to show that the GLS estimate of $\boldsymbol{\beta}$ when the block effects are random is the same as the GLS estimate of $\boldsymbol{\beta}$ when the block effects are fixed in the multiresponse case.

From the results given by Zellner(1963) we know that $\boldsymbol{\beta}$ given in equation (5.2.11) is the BLUE of $\boldsymbol{\beta}$. If we can show that the GLS estimate of $\boldsymbol{\beta}$ when the block effects are treated as fixed (when in fact they are random) is the same as $\boldsymbol{\beta}$ then the former estimator will have the same optimal properties as $\boldsymbol{\beta}$. This means that when using the GLS estimator in (4.5.12) we are not using an estimator that is less efficient than $\boldsymbol{\beta}$.

Suppose that the block effects are treated as fixed (when in fact they are random). Then, the variance-covariance matrix given in equation (5.2.8) simplifies to $\boldsymbol{\Sigma} \otimes \boldsymbol{I}_n$. Assuming that the design blocks orthogonally and using Theorem 4.5.1, the GLS estimate of $\boldsymbol{\beta}$ can be derived from

$$\hat{\boldsymbol{\psi}} = \begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta} \end{pmatrix} = [\boldsymbol{X}^{0\prime}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}^0]^{-1}\boldsymbol{X}^{0\prime}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} \qquad (5.3.1)$$

that is

$$\boldsymbol{\beta} = [\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X} - \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{X}]^{-1}[\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y} -$$

$$\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{y}| \tag{5 3 2}$$

where $\boldsymbol{V} = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n$. Equation (5 3 2) follows from equation (5 2 11) by replacing $\boldsymbol{\Phi}$ by $\boldsymbol{V}$. To show that $\hat{\boldsymbol{\beta}}$ (GLS estimator when block effects are random) given in equation (5 2 11) simplifies to $\hat{\boldsymbol{\beta}}$ (GLS estimator when block effects are fixed) given in equation (5 3 2) under orthogonal blocking we need to consider several identities which are established in the following lemmas

Lemma 5.3.1 *Let* $\boldsymbol{R} = \boldsymbol{\Sigma}_g \otimes \boldsymbol{I}_b$ *then*

$$\boldsymbol{\Phi}^{-1} = (\boldsymbol{V} + \boldsymbol{Z}\boldsymbol{R}\boldsymbol{Z}')^{-1} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{Z}(\boldsymbol{R}^{-1} + \boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}^{-1}, \tag{5 3 3}$$

and

$$\begin{aligned}(\boldsymbol{M}'\boldsymbol{\Phi}^{-1}\boldsymbol{M})^{-} &= (\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-} + (\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{Z}[(\boldsymbol{R}^{-1} + \boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z} - \\ &\quad \boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{Z}]^{-1}\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-} \\ &= (\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-} + \boldsymbol{C}\end{aligned} \tag{5 3 4}$$

where

$$\begin{aligned}\boldsymbol{C} &= (\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{Z}[\boldsymbol{R}^{-1} + \boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z} - \boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{Z}]^{-1} \\ &\quad \boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-}\end{aligned} \tag{5 3 5}$$

Proof The above results follow directly from Seber (1984, p 519, formula A3 3)

Lemma 5.3.2 *If a design blocks orthogonally then*

$$\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{Z} = \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z} \tag{5 3 6}$$

Proof Recall that $\boldsymbol{M} = \boldsymbol{I}_r \otimes \boldsymbol{1}_n$ and $\boldsymbol{V} = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n$. Then $(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-1} = \frac{1}{n}\boldsymbol{\Sigma}$. From these results we have

$$\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{Z} = \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{M}\frac{1}{n}\boldsymbol{\Sigma}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{Z}$$

$$= \boldsymbol{X}'\boldsymbol{V}^{-1}\frac{1}{n}(\boldsymbol{\Sigma} \otimes \boldsymbol{J}_n)\boldsymbol{V}^{-1}\boldsymbol{Z}$$

$$= \boldsymbol{X}'\boldsymbol{V}^{-1}\frac{1}{n}(\boldsymbol{\Sigma} \otimes \boldsymbol{J}_n)(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}$$

$$= \boldsymbol{X}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z} \qquad (5.3.7)$$

where $\boldsymbol{\mathcal{J}} = \boldsymbol{I}_r \otimes \boldsymbol{J}_n$. Using equation(4.5.9) we have $\boldsymbol{X}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z} = \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}$. Note that equation(4.5.9) is true provided that the design blocks orthogonally.

Lemma 5.3.3

$$\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{Z} = \boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z} \qquad (5.3.8)$$

Proof Using the results from Lemma 5.3.2 we have,

$$\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{Z} = \boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{M}\frac{1}{n}\boldsymbol{\Sigma}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{Z}$$

$$= \boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}(\boldsymbol{\Sigma} \otimes \boldsymbol{J}_n)\boldsymbol{V}^{-1}\boldsymbol{Z}$$

$$= \boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}(\boldsymbol{\Sigma} \otimes \boldsymbol{J}_n)(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}$$

$$= \boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z} \qquad (5.3.9)$$

Lemma 5.3.4 *Let* $\boldsymbol{\mathcal{D}} = -(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}$ *and* $\boldsymbol{E} = (\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z}-\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z})^{-1}$ *then*

$$\boldsymbol{\mathcal{D}}+\boldsymbol{E} = (\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z}-\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z})^{-1} \qquad (5.3.10)$$

Proof The above result follows from an identity given by Searle(1982 p 151 Exercise 16(e))

In Chapter 4 Theorem 4.5.1 we have derived an important result on the GLS estimator of β under orthogonal blocking when the block effects are considered to be zero. In the following theorem we demonstrate another advantage of using designs that block orthogonally. If the design used for a multiresponse experiment with random block effects blocks orthogonally then the GLS estimator of β is not affected by an incorrect analysis wherein the the block effects are (incorrectly) considered as fixed effects. Using Lemmas 5.3.1 to 5.3.4 we shall now prove that $\hat{\beta} = \beta$ in the following theorem

Theorem 5.3.1 If the design blocks orthogonally then the GLS estimate of β when the block effects are treated as fixed is equal to the GLS estimate of β when the block effects are random, that is $\hat{\beta} = \beta$

Proof Let

$$\mathcal{A} = \mathbf{X}'\Phi^{-1}\mathbf{X} - \mathbf{X}'\Phi^{-1}M(M'\Phi^{-1}M)^{-1}M'\Phi^{-1}\mathbf{X}, \tag{5.3.11}$$

$$\mathcal{B} = \mathbf{X}'\Phi^{-1}y - \mathbf{X}'\Phi^{-1}M(M'\Phi^{-1}M)^{-1}M'\Phi^{-1}y \tag{5.3.12}$$

then from equation (5.2.11) we have

$$\beta = \mathcal{A}^{-1}\mathcal{B} \tag{5.3.13}$$

Using equation (5.3.3) in Lemma 5.3.1 and equation (5.3.11), we can rewrite $\mathcal{A}$ as

$$\mathcal{A} = \mathcal{A}_1^* + \mathcal{A}_2^* \tag{5.3.14}$$

where

$$\begin{aligned}\mathcal{A}^* &= \mathbf{X}'\Phi^{-1}\mathbf{X} \\ &= \mathbf{X}'[V^{-1} - V^{-1}Z(R^{-1} + Z'V^{-1}Z)^{-1}Z'V^{-1}]\mathbf{X} \\ &= \mathcal{A}_1 + \mathcal{A}_2,\end{aligned} \tag{5.3.15}$$

$$\mathcal{A}_1 = X'V^{-1}X \tag{5 3 16}$$

$$\mathcal{A}_2 = -X'V^{-1}Z(R^{-1} + Z'V^{-1}Z)^{-1}Z'V^{-1}X \tag{5 3 17}$$

and

$$\mathcal{A}_3^* = -X\Phi^{-1}M(M\Phi^{-}M)^{-1}M'\Phi^{-1}X \tag{5 3 18}$$

$$= -X'[V^{-1} - V^{-1}Z(R^{-1} + Z'V^{-1}Z)^{-1}Z'V^{-1}]M\{M'[V^{-1} - V^{-1}Z(R^{-1} + Z'V^{-1}Z)^{-1}Z'V^{-1}]M\}^{-1}M'[V^{-1} - V^{-1}Z(R^{-1} + Z'V^{-1}Z)^{-1}Z'V^{-1}]X \tag{a 3 19}$$

From equation (5 3 4), we know that

$$(M'\Phi^{-1}M)^{-1} = \{M'[V^{-1} - V^{-1}Z(R^{-1} + Z'V^{-1}Z)^{-1}Z'V^{-1}]M\}^{-1}$$

$$= (M'V^{-1}M)^{-1} + C$$

where C is defined in equation (5 3 5) This implies that

$$\mathcal{A}_3^* = -X'[V^{-1} - V^{-1}Z(R^{-1} + Z'V^{-1}Z)^{-1}Z'V^{-1}]M[(M'V^{-1}M)^{-1} + C]M'[V^{-1} - V^{-1}Z(R^{-1} + Z'V^{-1}Z)^{-1}Z'V^{-1}]X$$

$$= \mathcal{A}_3 + \mathcal{A}_4 + \cdots + \mathcal{A}_{10} \tag{5 3 20}$$

where

$$\mathcal{A}_3 = -XV^{-1}M(M'V^{-1}M)^{-1}M'V^{-1}X \tag{5 3 21}$$

$$\mathcal{A}_4 = X'V^{-1}M(M'V^{-1}M)^{-1}M'V^{-1}Z(R^{-1} + Z'V^{-1}Z)^{-1}Z'V^{-1}X \tag{5 3 22}$$

$$\mathcal{A}_5 = -X'V^{-1}MCM'V^{-1}X \tag{5 3 23}$$

$$\mathcal{A}_6 = X'V^{-1}MCM'V^{-1}Z(R^{-1} + ZV^{-1}Z)^{-1}Z'V^{-1}X \tag{5 3 24}$$

$$\mathcal{A}_7 = XV^{-1}Z(R^{-1} + Z'V^{-1}Z)^{-1}Z'V^{-1}M(M'V^{-1}M)^{-1}M'V^{-1}X \tag{5 3 25}$$

$$\mathcal{A}_8 = -\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{Z}$$
$$(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{X} \quad (5.3.26)$$

$$\mathcal{A}_9 = \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{M}\boldsymbol{C}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{X} \quad (5.3.27)$$

$$\mathcal{A}_{10} = -\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{M}\boldsymbol{C}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{Z}(\boldsymbol{R}^{-1}+$$
$$\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{X} \quad (5.3.28)$$

To be able to rewrite $\hat{\boldsymbol{\beta}}$ in the desired form we apply Lemmas 5.3.2-5.3.3 to equations (5.3.22)-(5.3.28) to get the following results

$$\mathcal{A}_4 = \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}$$
$$\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{X}$$
$$= \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-}\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{X} \ (\textit{using equation}(5.3.6))$$
$$= -\mathcal{A}_2 \quad (5.3.29)$$

$$\mathcal{A}_5 = -\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{M}\{(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{Z}[(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})-\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{M}$$
$$(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{Z}]^{-1}\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-1}\}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{X}$$
$$(\textit{using equation}(5.3.5))$$
$$= -\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}[(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})-\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{Z}]^{-1}$$
$$\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{X} \ (\textit{using equation}(5.3.6))$$
$$= -\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}[(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})-\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{Z}]^{-1}$$
$$\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{X} \ (\textit{using equation}(5.3.6)) \quad (5.3.30)$$
$$= -\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}[(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})-\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{J}\boldsymbol{Z}]^{-1}\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{X}$$
$$(\textit{using equation}(5.3.8)) \quad (5.3.31)$$

Using the following result used in the simplification of $\mathcal{A}_5$ that is

$$\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{M}\boldsymbol{C} = \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}[(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})-\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})\ \boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{Z}]^{-1}$$
$$\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-1}$$

we can simplify $\mathcal{A}_6$ to the form

$$\begin{aligned}\mathcal{A}_6 &= \boldsymbol{X'V^{-1}Z[(R^{-1}+Z'V^{-}Z)-Z'V^{-1}M(M'V^{-1}M)^{-1}M'V^{-1}Z]^{-1}Z'V^{-1}M}\\ &\quad \boldsymbol{(M'V^{-1}M)^{-}M'V^{-1}Z(R^{-1}-Z'V^{-1}Z)^{-1}Z'V^{-1}X}\\ &= \boldsymbol{XV^{-1}Z[(R^{-1}+Z'V^{-1}Z)-Z'V^{-1}\frac{1}{n}\mathcal{J}Z]^{-1}Z'V^{-1}\frac{1}{n}\mathcal{J}Z(R^{-1}+Z'V^{-1}Z)^{-1}}\\ &\quad \boldsymbol{ZV^{-1}X}\ \textit{(using equation(5 3 8))} \qquad (5\ 3.32)\end{aligned}$$

From equations (5 3 25)-(5 3 28) we have

$$\mathcal{A}_7 = \boldsymbol{XV^{-1}Z(R^{-1}+Z'V^{-1}Z)^{-}Z'V^{-1}X}\ \textit{(using equation(5 3 6))} \qquad (5\ 3\ 33)$$

$$\mathcal{A}_8 = \boldsymbol{-X'V^{-1}Z(R^{-1}+Z'V^{-1}Z)^{-1}Z'V^{-1}\frac{1}{n}\mathcal{J}Z(R^{-1}+Z'V^{-1}Z)^{-1}}$$
$$\boldsymbol{Z'V^{-1}X}\ \textit{(using equation(5 3 8))} \qquad (5\ 3\ 34)$$

$$\begin{aligned}\mathcal{A}_9 &= \boldsymbol{X'V^{-1}Z(R^{-1}+Z'V^{-1}Z)^{-1}Z'V^{-1}M\{(M'V^{-1}M)^{-1}M'V^{-1}Z[R^{-1}+}\\ &\quad \boldsymbol{Z'V^{-1}Z-Z'V^{-1}M(M'V^{-1}M)^{-1}M'V^{-1}Z]^{-1}Z'V^{-1}M(M'V^{-1}M)^{-1}\}}\\ &\quad \boldsymbol{M'V^{-1}X}\ \textit{(using equation(5 3 5))}\\ &= \boldsymbol{XV^{-1}Z(R^{-1}+Z'V^{-1}Z)^{-1}Z'V^{-1}\frac{1}{n}\mathcal{J}Z[(R^{-1}+Z'V^{-1}Z)-}\\ &\quad \boldsymbol{Z'V^{-1}\frac{1}{n}\mathcal{J}Z]^{-1}Z'V^{-1}M(M'V^{-1}M)^{-1}M'V^{-1}X}\ \textit{(using equation(5 3 8))}\\ &= \boldsymbol{X'V^{-1}Z(R^{-1}+Z'V^{-1}Z)^{-1}Z'V^{-1}\frac{1}{n}\mathcal{J}Z[(R^{-1}+Z'V^{-1}Z)-}\\ &\quad \boldsymbol{ZV^{-1}\frac{1}{n}\mathcal{J}Z]^{-1}Z'V^{-1}X}\ \textit{(using equation(5 3 6))} \qquad (5\ 3\ 35)\end{aligned}$$

$$\begin{aligned}\mathcal{A}_{10} &= \boldsymbol{-X'V^{-1}Z(R^{-1}+Z'V^{-1}Z)^{-1}Z'V^{-1}M\{(M'V^{-1}M)^{-1}M'V^{-1}Z}\\ &\quad \boldsymbol{[(R^{-1}+Z'V^{-1}Z)-ZV^{-1}M(M'V^{-1}M)^{-1}M'V^{-1}Z]^{-1}Z'V^{-1}M}\\ &\quad \boldsymbol{(M'V^{-1}M)^{-}\}MV^{-1}Z(R^{-1}+Z'V^{-1}Z)^{-1}Z'V^{-1}X}\\ &\quad \textit{(using equation(5 3 5))}\end{aligned}$$

$$= -\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}[(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})-$$

$$\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}]^{-1}\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{X}$$

$$(using\ equation (5\ 3\ 8)) \qquad (5\ 3\ 36)$$

Consider equations (5 3 31) - (5 3 36) for the terms $\boldsymbol{\mathcal{A}}_6$ to $\boldsymbol{\mathcal{A}}_{10}$. We can further simplify $\boldsymbol{\mathcal{A}}_5^*$ by looking at the following groupings of these terms

$$\boldsymbol{\mathcal{A}}_5+\boldsymbol{\mathcal{A}}_6 = -\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}[(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})-\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}]^{-1}$$

$$[\boldsymbol{I}_b-\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}]\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{X} \qquad (5\ 3\ 37)$$

$$\boldsymbol{\mathcal{A}}_7+\boldsymbol{\mathcal{A}}_8 = \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}[\boldsymbol{I}_b-\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}(\boldsymbol{R}^{-1}+$$

$$\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}]\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{X} \qquad (5\ 3\ 38)$$

$$\boldsymbol{\mathcal{A}}_9+\boldsymbol{\mathcal{A}}_{10} = \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}[(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})-$$

$$\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}]^{-1}[\boldsymbol{I}_b-\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}]$$

$$\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{X} \qquad (5\ 3\ 39)$$

Thus using equations (5 3 14) (5.3 17), (5 3 20) (5 3 21), (5 3 29) and (5 3 37) - (5 3 39), we have

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{A}}_1+\boldsymbol{\mathcal{A}}_2+\boldsymbol{\mathcal{A}}_3-\boldsymbol{\mathcal{A}}_4+(\boldsymbol{\mathcal{A}}_5+\boldsymbol{\mathcal{A}}_6)+(\boldsymbol{\mathcal{A}}_7+\boldsymbol{\mathcal{A}}_8)+(\boldsymbol{\mathcal{A}}_9+\boldsymbol{\mathcal{A}}_{10})$$

$$= \boldsymbol{\mathcal{A}}_1+\boldsymbol{\mathcal{A}}_3+(\boldsymbol{\mathcal{A}}_5+\boldsymbol{\mathcal{A}}_6)+(\boldsymbol{\mathcal{A}}_7+\boldsymbol{\mathcal{A}}_8)+(\boldsymbol{\mathcal{A}}_9+\boldsymbol{\mathcal{A}}_{10})$$

that is

$$\boldsymbol{\mathcal{A}} = \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}-\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{X}$$

$$-\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}[(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})-\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}]^{-1}[\boldsymbol{I}_b-$$

$$
\begin{aligned}
& \boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}]\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{X} \\
+ \; & \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}[\boldsymbol{I}_b-\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}(\boldsymbol{R}^{-} - \\
& \boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}]\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{X} \\
+ \; & \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}\{\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}[(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})- \\
& \boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}]^{-1}[\boldsymbol{I}_b-\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}] \\
& \boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{X} \qquad (5.3.41)
\end{aligned}
$$

Combining and grouping similar terms we have

$$
\begin{aligned}
\boldsymbol{\mathcal{A}} \; = \; & \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}-\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{X} \\
+ \; & \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}\{-[(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})-\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}]^{-1}+(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1} \\
+ \; & (\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}[(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})-\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}]^{-1}\} \\
& [\boldsymbol{I}_b-\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}]\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{X} \qquad (5.3.41)
\end{aligned}
$$

Similarly we can rewrite $\boldsymbol{\mathcal{B}}$ as follows

$$
\begin{aligned}
\boldsymbol{\mathcal{B}} \; = \; & \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}-\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-}\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{y} \\
+ \; & \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}\{-[(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})-\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}]^{-1}+(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1} \\
+ \; & (\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}[(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})-\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}]^{-1}\} \\
& [\boldsymbol{I}_b-\boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}]\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{y} \qquad (5.3.42)
\end{aligned}
$$

Applying Lemma 5.3.4 to equation (5.3.41) we have

$$
\boldsymbol{\mathcal{A}} \; = \; \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}-\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{X}+
$$

$$\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}[-\boldsymbol{E}-\boldsymbol{\mathcal{D}}+(\boldsymbol{\mathcal{D}}+\boldsymbol{E})][\boldsymbol{I}_b -$$

$$\boldsymbol{Z}\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}]\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{X} \qquad (5\,3\,43)$$

Similarly,

$$\begin{aligned}\boldsymbol{B} &= \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}-\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{y}+\\ &\quad \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Z}[-\boldsymbol{E}-\boldsymbol{\mathcal{D}}+(\boldsymbol{\mathcal{D}}+\boldsymbol{E})][\boldsymbol{I}_b -\\ &\quad \boldsymbol{Z}'\boldsymbol{V}^{-1}\frac{1}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}(\boldsymbol{R}^{-1}+\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{Z})^{-1}]\boldsymbol{Z}'\boldsymbol{V}^{-1}\boldsymbol{y} \qquad (5\,3\,44)\end{aligned}$$

Thus from equations (5.3 43) (5 3 44), and (5 3.13) it follows that

$$\begin{aligned}\boldsymbol{\beta} &= [\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}-\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{X}]^{-1}[\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}-\\ &\quad \boldsymbol{X}\boldsymbol{V}^{-1}\boldsymbol{M}(\boldsymbol{M}\boldsymbol{V}^{-1}\boldsymbol{M})^{-1}\boldsymbol{M}'\boldsymbol{V}^{-1}\boldsymbol{y}] \qquad (5\,3\,45)\\ &= \boldsymbol{\beta} \qquad (5\,3\,46)\end{aligned}$$

5.4 Estimating the Parameters in Multiresponse Models with Mixed Effects

The interactions between the fixed polynomial effects and the random effects $\boldsymbol{\delta}_1$ $\boldsymbol{\delta}_2$, , $\boldsymbol{\delta}_r$ are ignored in model (5 2 1) In this section these interaction effects are added to the model This introduces additional random effects into the models In this case, the resulting process variance contains terms for the block effects and the polynomial interaction terms The interaction terms are essentially the products of corresponding polynomial terms (elements of $\boldsymbol{X}_i$) and block effect terms (elements of $\boldsymbol{Z}_i$) Furthermore since the $\boldsymbol{Z}_i$s consist of ones and zeros only, the values of these products basically depend on the elements of $\boldsymbol{X}_i$ (which in turn depend on the settings used for the input variables) As a result the process variance will depend on the settings of the input variables In addition since the process variance contains

interaction terms then it follows that these interactions can affect the determination of optimum conditions. Hence the addition of interaction terms between the fixed polynomial effects and the random effects $\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \ldots, \boldsymbol{\delta}_r$ should be considered. The model for the ith response with the added interaction terms is given by

$$\boldsymbol{y}_i = \beta_{0i}\mathbf{1}_n + \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{Z}_i\boldsymbol{\delta}_i + \sum_{k=1}^{p} \boldsymbol{T}_{ik}\boldsymbol{\Delta}_{ik} + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \ldots, r, \tag{5.4.1}$$

where $\beta_{0i}, \boldsymbol{X}_i, \boldsymbol{\beta}_i, \boldsymbol{Z}_i, \boldsymbol{\delta}_i$ and $\boldsymbol{\varepsilon}_i$ are as defined in model (5.2.1), $\boldsymbol{T}_{ik}$ is a matrix of order $n \times b$ whose lth column is obtained by multiplying the elements of the kth column of $\boldsymbol{X}_i$ with the corresponding elements of the lth column of $\boldsymbol{Z}_i$, $k = 1, 2, \ldots, p_i$, $l = 1, 2, \ldots, b$, and $\boldsymbol{\Delta}_{ik}$ is a vector of interaction coefficients of order $b \times 1$ associated with the kth column of $\boldsymbol{X}_i$, $k = 1, 2, \ldots, p_i$. Model (5.4.1) can be rewritten as

$$\boldsymbol{y}_i = \beta_{0i}\mathbf{1}_n + \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{w}_i, \quad i = 1, 2, \ldots, r \tag{5.4.2}$$

or equivalently

$$\boldsymbol{y}_i = \boldsymbol{U}_i\boldsymbol{\Xi}_i + \boldsymbol{w}_i, \quad i = 1, 2, \ldots, r, \tag{5.4.3}$$

where

$$\boldsymbol{U}_i = [\mathbf{1}_n \quad \boldsymbol{X}_i], \tag{5.4.4}$$

$$\boldsymbol{\Xi}_i = (\beta_{0i}, \boldsymbol{\beta}_i')' \tag{5.4.5}$$

$$\boldsymbol{w}_i = \boldsymbol{Z}_i\boldsymbol{\delta}_i + \boldsymbol{T}_i^*\boldsymbol{\Delta}_i^* + \boldsymbol{\varepsilon}_i, \tag{5.4.6}$$

$$\boldsymbol{T}_i^* = [\boldsymbol{T}_{i1} \quad \boldsymbol{T}_{i2} \quad \ldots \quad \boldsymbol{T}_{ip_i}], \tag{5.4.7}$$

$$\boldsymbol{\Delta}_i^* = (\boldsymbol{\Delta}_{i1}' \quad \boldsymbol{\Delta}_{i2}' \quad \ldots \quad \boldsymbol{\Delta}_{ip_i}')' \tag{5.4.8}$$

Since the block effects are random, then $\boldsymbol{\delta}_i$ and $\boldsymbol{\Delta}_i^*$ are random vectors ($i = 1, 2, \ldots, r$). Assume the following:

- $\boldsymbol{\Delta}_{ik} \sim N(0, \sigma_{ik}^{(1)}\boldsymbol{I}_b)$, $i = 1, 2, \ldots, r$, $k = 1, 2, \ldots, p_i$,

- $E(\boldsymbol{\Delta}_{ik}\boldsymbol{\Delta}_{jl}') = \begin{cases} \sigma_{kl}^{ij}\boldsymbol{I}_{p_i} & i \neq j,\ i,j = 1,2,\ldots,r \quad k,l = 1,2,\ldots,p_i \\ \mathbf{0} & i = j,\ i,j = 1,2,\ldots,r \quad k \neq l \\ \sigma_{kk}^{ij}\boldsymbol{I}_{p_i} & i = j,\ i,j = 1,2,\ldots,r \quad k = l \end{cases}$

  hence $E(\boldsymbol{\Delta}_i^*\boldsymbol{\Delta}_j^{*\prime}) = \boldsymbol{\Sigma}^{ij} \otimes \boldsymbol{I}_b$, where $\boldsymbol{\Sigma}^{ij}$ is a matrix of order $p_i \times p_j$ given by $\boldsymbol{\Sigma}^{ij} = (\sigma_{kl}^{ij})$ $i \neq j,\ i,j = 1,2,\ldots,r$ $k = 1,2,\ldots,p_i$ $l = 1,2,\ldots,p_j$ and $E(\boldsymbol{\Delta}_i^*\boldsymbol{\Delta}_i^{*\prime}) = \boldsymbol{\Sigma}^{ii} \otimes \boldsymbol{I}_b$, where $\boldsymbol{\Sigma}^{ii} = \mathrm{diag}(\sigma_{11}^{ii}, \sigma_{22}^{ii}, \ldots, \sigma_{p_ip_i}^{ii})$, $i = 1,2,\ldots,r$

- $\boldsymbol{\delta}_i \sim N(\mathbf{0}, \sigma_{\delta ii}\boldsymbol{I}_b)$ $i = 1,2,\ldots,r$

- $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma_{ii}\boldsymbol{I}_n)$ $i = 1,2,\ldots,r$

- $\boldsymbol{\Delta}_i^*$, $\boldsymbol{\delta}_i$ and $\boldsymbol{\epsilon}_i$ are independent for a fixed i $i = 1,2,\ldots,r$,

- $\boldsymbol{\Delta}_i^*$ and $\boldsymbol{\delta}_j$ are independent for $i \neq j$, $i,j = 1,2,\ldots,r$,

- $\boldsymbol{\Delta}_i^*$ and $\boldsymbol{\epsilon}_j$ are independent for $i \neq j$ $i,j = 1,2,\ldots,r$

- $\boldsymbol{\delta}_i$ and $\boldsymbol{\epsilon}_j$ are independent for $i \neq j$ $i,j = 1,2,\ldots,r$

- $E(\boldsymbol{\delta}_i\boldsymbol{\delta}_j') = \sigma_{\delta ij}\boldsymbol{I}_b$, $i \neq j$, $i,j = 1,2,\ldots,r$ and

- $E(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_j') = \sigma_{ij}\boldsymbol{I}_n$, $i \neq j$ $i,j = 1,2,\ldots,r$

Let $\boldsymbol{\Gamma}_{ij} = E(\boldsymbol{w}_i\boldsymbol{w}_j')$ then these assumptions imply that

$$\boldsymbol{\Gamma}_{ij} = \sigma_{\delta ij}\boldsymbol{Z}_i\boldsymbol{Z}_j' + \boldsymbol{T}_i^*(\boldsymbol{\Sigma}^{ij} \otimes \boldsymbol{I}_b)\boldsymbol{T}_j^{*\prime} + \sigma_{ij}\boldsymbol{I}_n, \quad i \neq j,\ i,j = 1,2,\ldots,r$$

$$\boldsymbol{\Gamma}_{ii} = \sigma_{\delta ii}\boldsymbol{Z}_i\boldsymbol{Z}_i' + \boldsymbol{T}_i^*(\boldsymbol{\Sigma}^{ii} \otimes \boldsymbol{I}_b)\boldsymbol{T}_i^{*\prime} + \sigma_{ii}\boldsymbol{I}_n, \quad i = 1,2,\ldots,r \tag{5.4.9}$$

The multiresponse form of the models given in equation (5.4.3) is

$$\boldsymbol{y} = \boldsymbol{U}\boldsymbol{\Xi} + \boldsymbol{w} \tag{5.4.10}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1', \boldsymbol{y}_2', \ldots, \boldsymbol{y}_r')'$ $\boldsymbol{U} = \mathrm{diag}(\boldsymbol{U}_1, \boldsymbol{U}_2, \ldots, \boldsymbol{U}_r)$ $\boldsymbol{\Xi} = (\boldsymbol{\Xi}_1', \boldsymbol{\Xi}_2', \ldots, \boldsymbol{\Xi}_r')'$ and $\boldsymbol{w} = (\boldsymbol{w}_1', \boldsymbol{w}_2', \ldots, \boldsymbol{w}_r')'$ We have

$$E(\boldsymbol{w}) = \mathbf{0} \tag{5.4.11}$$

and from (5.4.9)

$$E(\boldsymbol{u}\boldsymbol{u}') = \boldsymbol{Z}(\boldsymbol{\Sigma}_{\delta} \otimes \boldsymbol{I}_{\delta})\boldsymbol{Z}' + \boldsymbol{T}^{*}(\boldsymbol{\Sigma}_{\Delta} \otimes \boldsymbol{I}_{\delta})\boldsymbol{T}^{*} + (\boldsymbol{\Sigma} \otimes \boldsymbol{I}_{n}) = \boldsymbol{\Gamma} \qquad (5.4.12)$$

where $\boldsymbol{Z} = diag(\boldsymbol{Z}_1 \ \boldsymbol{Z}_2 \quad \boldsymbol{Z}_r)$, $\boldsymbol{T}^{*} = diag(\boldsymbol{T}_1^{*}, \boldsymbol{T}_2^{*}, \ldots, \boldsymbol{T}_r^{*})$, $\boldsymbol{\Sigma}_{\delta} = (\sigma_{\delta\eta})$ and $\boldsymbol{\Sigma}_{\Delta}$ is the partitioned matrix $\boldsymbol{\Sigma}_{\Delta} = [\boldsymbol{\Sigma}^{ij}]$. The GLS estimator of $\boldsymbol{\Xi}$ is given by

$$\boldsymbol{\Xi} = (\boldsymbol{U}'\boldsymbol{\Gamma}^{-1}\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{\Gamma}^{-1}\boldsymbol{y} \qquad (5.4.13)$$

and its variance-covariance matrix is

$$Var(\boldsymbol{\Xi}) = (\boldsymbol{U}'\boldsymbol{\Gamma}^{-1}\boldsymbol{U})^{-1} \qquad (5.4.14)$$

The variance-covariance matrix $\boldsymbol{\Gamma}$ is usually unknown and therefore must be estimated in order to obtain the solutions in (5.4.13) and (5.4.14). A consistent estimator of $\boldsymbol{\Gamma}$ can be obtained by following the procedure described by Singh and Ullah (1974, pp. 192-193). This procedure is essentially a multiresponse analogue of the Hildreth and Houck (1968) procedure for single-response models with random coefficients. The procedure involves the establishment of a regression relation between the ordinary least squares residuals obtained from each fitted ith response model and a linear combination of the matrices given by $\boldsymbol{U}_i \ \boldsymbol{Z}_{i1}, \boldsymbol{T}_i^{*}$ and $\boldsymbol{I}_n - \boldsymbol{U}_i(\boldsymbol{U}_i'\boldsymbol{U}_i)^{-1}\boldsymbol{U}_i'$, $i = 1, 2, \ldots, r$, which are used to obtain these residuals. In this regression relation the parameters of interest are the variances of the random effects in the model. We illustrate this procedure for the multiresponse case by following the steps outlined by Raj and Ullah (1981, pp. 241-257). First, we obtain the ordinary least squares (OLS) residuals for each equation in (5.4.3) and write

$$\boldsymbol{w}_i = \boldsymbol{M}_i\boldsymbol{y}_i, \quad i = 1, 2, \ldots, r \qquad (5.4.15)$$

where $\boldsymbol{M}_i = \boldsymbol{I}_n - \boldsymbol{U}_i(\boldsymbol{U}_i'\boldsymbol{U}_i)^{-1}\boldsymbol{U}_i'$. Since $\boldsymbol{M}_i\boldsymbol{U}_i = \boldsymbol{O}$ then

$$\boldsymbol{M}_i\boldsymbol{w}_i = \boldsymbol{M}_i(\boldsymbol{y}_i - \boldsymbol{U}_i\boldsymbol{\Xi}_i) = \boldsymbol{M}_i\boldsymbol{y}_i, \quad i = 1, 2, \ldots, r$$

It follows that

$$\boldsymbol{w}_i = \boldsymbol{M}_i \boldsymbol{w}, \tag{5.4.16}$$

Further, using (5.4.9) we obtain

$$E(\boldsymbol{w}_i \boldsymbol{w}_j') = \begin{cases} \boldsymbol{M}_i[\sigma_{kj}\boldsymbol{Z}_1\boldsymbol{Z}_j' + \boldsymbol{T}_i^*(\boldsymbol{\Sigma}^{ij} \otimes \boldsymbol{I}_b)\boldsymbol{T}_j^* + \sigma_{ij}\boldsymbol{I}_n]\boldsymbol{M}_j, & i \neq j \\ \boldsymbol{M}_i[\sigma_{kk}\boldsymbol{Z}_1\boldsymbol{Z}_i' + \boldsymbol{T}_i^*(\boldsymbol{\Sigma}^{ii} \otimes \boldsymbol{I}_b)\boldsymbol{T}_i^* + \sigma_{ii}\boldsymbol{I}_n]\boldsymbol{M}_i, & i = j \end{cases} \tag{5.4.17}$$

We can rewrite equation (5.4.17) in the form

$$E(\boldsymbol{w}_i \boldsymbol{w}_j') = \begin{cases} \boldsymbol{M}_i[\sigma_{kj}\boldsymbol{Z}_1\boldsymbol{Z}_j' + \sum_{k=1}^{p_i}\sum_{l=1}^{p}\boldsymbol{T}_{ik}\boldsymbol{T}_{jl}'\sigma_{kl}^{ij} + \boldsymbol{I}_n\sigma_{ij}]\boldsymbol{M}_j & i \neq j \\ \boldsymbol{M}_i[\sigma_{kk}\boldsymbol{Z}_1\boldsymbol{Z}_i' + \sum_{k=1}^{p_i}\boldsymbol{T}_{ik}\boldsymbol{T}_{ik}'\sigma_{kk}^{ii} + \boldsymbol{I}_n\sigma_{ii}]\boldsymbol{M}_i & i = j \end{cases} \tag{5.4.18}$$

that is

$$E(\boldsymbol{w}_i \boldsymbol{w}_j') = \begin{cases} \boldsymbol{M}_i(\boldsymbol{Z}_1\boldsymbol{Z}_j'\boldsymbol{M}_j\sigma_{kj} + \boldsymbol{T}_{i1}\boldsymbol{T}_{j1}'\boldsymbol{M}_j\sigma_{11}^{ij} + \quad + \boldsymbol{T}_{i1}\boldsymbol{T}_{jp_j}'\boldsymbol{M}_j\sigma_{1p_j}^{ij} + \\ +\boldsymbol{T}_{ip_i}\boldsymbol{T}_{j1}'\boldsymbol{M}_j\sigma_{p_i1}^{ij} + \boldsymbol{M}_j\sigma_{ij}) & i \neq j, \\ \boldsymbol{M}_i(\boldsymbol{Z}_1\boldsymbol{Z}_i'\boldsymbol{M}_i\sigma_{kk} + \boldsymbol{T}_{i1}\boldsymbol{T}_{i1}'\boldsymbol{M}_i\sigma_{11}^{ii} + \boldsymbol{T}_{i2}\boldsymbol{T}_{i2}'\boldsymbol{M}_i\sigma_{22}^{ii} + \\ +\boldsymbol{T}_{ip_i}\boldsymbol{T}_{ip}'\boldsymbol{M}_i\sigma_{p\,p_i}^{ii} + \boldsymbol{M}_i\sigma_{ii}), & i = j \end{cases} \tag{5.4.19}$$

The $(c\ d)$th element of $E(\boldsymbol{w}_i\boldsymbol{w}_j')$ is given by

$$E(u_{ic}u_{jd}) = \begin{cases} \boldsymbol{m}_{cd}^{ij}\boldsymbol{\sigma}_{ij}^* & i \neq j \quad c,d = 1\ 2 \quad\quad n \\ \boldsymbol{m}_{cd}^{ii}\boldsymbol{\sigma}_{ii}^* & i = j \quad c,d = 1\ 2 \quad\quad n \end{cases} \tag{5.4.20}$$

where $\boldsymbol{m}_{cd}^{ij}$ $i\ j = 1\ 2 \quad\quad r$ is given by

$$\begin{aligned} \boldsymbol{m}_{cd}^{ij} \;=\; & [\boldsymbol{m}_{ic}'\boldsymbol{Z}_1\boldsymbol{Z}_j'\boldsymbol{m}_{jd} \quad \boldsymbol{m}_{ic}'\boldsymbol{T}_{i1}\boldsymbol{T}_{j1}'\boldsymbol{m}_{jd} \quad\quad \boldsymbol{m}_{ic}'\boldsymbol{T}_{i1}\boldsymbol{T}_{jp_j}'\boldsymbol{m}_{jd} \\ & \boldsymbol{m}_{ic}'\boldsymbol{T}_{ip_i}\boldsymbol{T}_{jp_j}'\boldsymbol{m}_{jd} \quad \boldsymbol{m}_{ic}'\boldsymbol{I}_n\boldsymbol{m}_{jd}], \end{aligned} \tag{5.4.21}$$

where $\boldsymbol{m}_{ic}'$ and $\boldsymbol{m}_{id}'$ are row vectors of order $1 \times n$ which consist of the elements in the cth and dth rows of $\boldsymbol{M}_i$, respectively $\boldsymbol{m}_{jd}'$ is also a row vector of order $1 \times n$ which consists of the elements in the dth row of $\boldsymbol{M}_j$ and

$$\boldsymbol{\sigma}_{ij}^* = \begin{cases} (\sigma_{kj}, \sigma_{11}^{ij}\ \sigma_{12}^{ij}, & \sigma_{1p_j}^{ij}, \sigma_{21}^{ij} \quad\quad, \sigma_{2p_j}^{ij} \quad\quad, -\sigma_{p_i1}^{ij}, \sigma_{ij})' & i \neq j \\ (\sigma_{kk}, \sigma_{11}^{ii}, \sigma_{22}^{ii} & \sigma_{p_ip_i-1}^{ij})' & i = j \end{cases} \tag{5.4.22}$$

Define

$$w_{ij} = vec(w_i w_j) \quad i, j = 1, 2, \ldots, r$$

where $vec(\,)$ is a matrix operator (see Searle(1982, pp 332-333)) which stacks the columns of the matrix inside the parentheses one under the other to form a single column. Then from equation (5.4.20) we can write

$$E(w_{ij}) = \begin{cases} \mathcal{M}_{ij}\sigma^*_{ij} & i \neq j,\ i, j = 1, 2, \ldots, r, \\ \mathcal{M}_{ii}\sigma^*_{ii} & i = j,\ i = 1, 2, \ldots, r, \end{cases} \tag{5.4.23}$$

where $\mathcal{M}_{ij}$ $(i \neq j)$ is a matrix of order $n^2 \times (p_i p_j + 2)$ of rank $p_i p_j + 2$ given by

$$\mathcal{M}_{ij} = [(m^{ij}_{11})' \quad (m^{ij}_{21})' \quad \ldots \quad (m^{ij}_{n1})' \quad \ldots \quad (m^{ij}_{nn})']', \quad i \neq j, \tag{5.4.24}$$

and $\mathcal{M}_{ii}$ is a matrix of order $n^2 \times (p_i + 2)$ of rank $p_i + 2$, which can also be written as

$$\mathcal{M}_{ii} = [(m^{ii}_{11})' \quad (m^{ii}_{12})' \quad \ldots \quad (m^{ii}_{n1})' \quad \ldots \quad (m^{ii}_{nn})']', \quad i = 1, 2, \ldots, r \tag{5.4.25}$$

Further, if we write

$$e_{ij} = \begin{cases} w_{ij} - E(w_{ij}) & i \neq j \\ w_{ii} - E(w_{ii}) & i = j, \end{cases}$$

then the regression relation implied by (5.4.23) can be written as

$$w_{ij} = \begin{cases} \mathcal{M}_{ij}\sigma^*_{ij} + e_{ij}, & i \neq j, i, j = 1, 2, \ldots, r \\ \mathcal{M}_{ii}\sigma^*_{ii} + e_{ii} & i = j, i = 1, 2, \ldots, r \end{cases} \tag{5.4.26}$$

If ordinary least squares (OLS) is used to estimate σ^*_{ij} in (5.4.26) we obtain a multiresponse analogue of one of the Hildreth and Houck(1968) estimators. This estimator is given by

$$\sigma^*_{ij} = \begin{cases} (\mathcal{M}'_{ij}\mathcal{M}_{ij})^- \mathcal{M}'_{ij} w_{ij} & i \neq j \\ (\mathcal{M}'_{ii}\mathcal{M}_{ii})^{-1}\mathcal{M}'_{ii} w_{ii} & i = j \end{cases} \tag{5.4.27}$$

Once all the σ^*_{ij}'s ($i, j = 1, 2, \ldots, r$) are obtained, we can determine the estimates of the matrices $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}_\delta$ and $\boldsymbol{\Sigma}_\Delta$. Substituting these estimated values in equation (5.4.12) we have

$$\boldsymbol{\Gamma} = \boldsymbol{Z}(\boldsymbol{\Sigma}_\delta \otimes \boldsymbol{I}_b)\boldsymbol{Z}' + \boldsymbol{T}^*(\boldsymbol{\Sigma}_\Delta \otimes \boldsymbol{I}_b)\boldsymbol{T}^{*\prime} + (\boldsymbol{\Sigma} \otimes \boldsymbol{I}_n) \tag{5.4.28}$$

Note that other estimators of $\boldsymbol{\Gamma}$ can be obtained. For example, GLS estimation may be applied to equation (5.4.26). Other approaches suggested by Srivastava and Giles (1987, pp. 336-337) include obtaining σ^*_{ij} as a minimum norm quadratic estimator.

Using $\boldsymbol{\Gamma}$ in place of $\boldsymbol{\Gamma}$ in equation (5.4.13), an EGLS estimator of $\boldsymbol{\Xi}$ is given by

$$\boldsymbol{\Xi}_e = (\boldsymbol{U}'\boldsymbol{\Gamma}^{-1}\boldsymbol{U})^{-1}\boldsymbol{U}'\boldsymbol{\Gamma}^{-1}\boldsymbol{y} \tag{5.4.29}$$

Singh and Ullah (1974, p. 193) showed that $\boldsymbol{\Xi}_e$ is consistent and that the asymptotic distribution of $n^{\frac{1}{2}}(\boldsymbol{\Xi}_e - \boldsymbol{\Xi})$ is normal. Hence an approximate expression for the variance-covariance matrix of $\boldsymbol{\Xi}_e$ is given by

$$Var(\boldsymbol{\Xi}_e) = (\boldsymbol{U}'\boldsymbol{\Gamma}^{-1}\boldsymbol{U})^{-1} \tag{5.4.30}$$

5.5 Mean Response and Prediction Variance

The mean of the ith response at a point $\boldsymbol{x} = (x_1, x_2, \ldots, x_k)'$ inside the experimental region $\mathfrak{R}$ is given by

$$\eta_i(\boldsymbol{x}) = \boldsymbol{\zeta}_i'(\boldsymbol{x})\boldsymbol{\Xi}, \quad i = 1, 2, \ldots, r, \tag{5.5.1}$$

where $\boldsymbol{\zeta}_i'(\boldsymbol{x}) = [1, \boldsymbol{f}_1'(\boldsymbol{x}), \ldots, \boldsymbol{f}_r'(\boldsymbol{x})]$, $\boldsymbol{f}_i'(\boldsymbol{x})$ is a vector of the same form as a row of $\boldsymbol{X}$, in (5.4.2), but is evaluated at the point $\boldsymbol{x}$, $i = 1, 2, \ldots, r$. From (5.5.1) and using the results from Section 4.4, the predicted value for the mean of the ith response, $i = 1, 2, \ldots, r$, at a point $\boldsymbol{x} = (x_1, x_2, \ldots, x_k)'$ in $\mathfrak{R}$ is of the form

$$y_{ei}(\boldsymbol{x}) = \boldsymbol{\zeta}_i'(\boldsymbol{x})\boldsymbol{\Xi}_e, \quad i = 1, 2, \ldots, r \tag{5.5.2}$$

where Ξ_r is obtained from equation (5.4.29). The multiresponse form of equation (5.5.2) is given by

$$\hat{y}_r(x) = \mathcal{G}'(x)\Xi_r \tag{5.5.3}$$

where $\mathcal{G}(x) = diag(\xi_1(x)\ \xi_2(x) \quad \xi_r(x))$. The estimated prediction variance matrix is approximately equal to

$$Var[\hat{y}_r(x)] = \mathcal{G}'(x)(U'\Gamma^{-1}U)^{-1}\mathcal{G}(x) \tag{5.5.4}$$

5.6 Optimization

To determine the optima of the r responses of interest, we use the multiresponse optimization procedure described in Section 3.3. The simultaneous optima of the r responses can be determined by finding conditions on x that minimize the distance function given by

$$\rho[\hat{y}_r(x)\ \phi] = \{[\hat{y}_r(x) - \phi]'[\mathcal{G}'(x)(U'\Gamma^{-1}U)^{-1}\mathcal{G}(x)]^{-1}[\hat{y}_r(x) - \phi]\}^{\frac{1}{2}} \tag{5.6.1}$$

where the elements of ϕ are the individual optima of the elements of $\hat{y}_r(x)$ over the experimental region $\Re$.

The individual optima and the minimum value of $\rho[\hat{y}_r(x), \phi]$ can be obtained by using the controlled random search procedure presented by Price (1977). A Fortran program that implements the procedure has been written by Conlon (1992). A description of Price's procedure as implemented by Conlon's Fortran program (first mentioned in Chapter 2) is given in Appendix B, Section B.3.

The steps to be taken to find the minimum of the distance function given in equation (5.6.1) are as follows:

1. Estimates of Σ, Σ_Δ and Σ_Δ are obtained according to equation (5.4.27) to determine the estimate of Γ from (5.4.28).

2 The EGLS estimator of Ξ is then obtained using (5 4 29)

3 The estimated responses are optimized individually to obtain the vector ϕ of estimated individual optima. The controlled random search procedure given by Price can be used to obtain these individual optima.

4 The distance function $\rho[\boldsymbol{y}_e(\boldsymbol{x}), \boldsymbol{\phi}]$ is minimized over $\mathbb{R}$. Again, Price's procedure can be used to obtain this minimum.

5.7 Closeness to Target Values

In some experimental situations it may be desired to determine the settings of the independent variables that result in a predicted response close to a target value t. The simplest way to measure the closeness of the predicted responses to the target values is to use the procedure described in Section 2.3, that is, we use a distance function. Let t_i, $i = 1, 2, \ldots, r$ be the target value for the ith response and let $\boldsymbol{t} = (t_1, t_2, \ldots, t_r)$. We measure the closeness of $\boldsymbol{y}_e(\boldsymbol{x})$ to $\boldsymbol{t}$ by using the distance function given in equation (5.6.1), with ϕ replaced by $\boldsymbol{t}$. The distance function in this case is given by

$$\rho[\boldsymbol{y}_e(\boldsymbol{x}), \boldsymbol{t}] = [(\boldsymbol{y}_e(\boldsymbol{x}) - \boldsymbol{t})'\{\boldsymbol{G}'(\boldsymbol{x})(\boldsymbol{U}'\boldsymbol{\Gamma}^{-1}\boldsymbol{U})^{-1}\boldsymbol{G}(\boldsymbol{x})\}^{-1}(\boldsymbol{y}_e(\boldsymbol{x}) - \boldsymbol{t})]^{\frac{1}{2}} \qquad (5.7.1)$$

Thus in order to achieve a certain degree of closeness to $\boldsymbol{t}$, values of $\boldsymbol{x}$ should be determined such that the distance function given in (5.7.1) is minimized.

5.8 A Numerical Example

For this example, we use the results of the experiment by Chen et al. (1989) mentioned in Section 4.6 concerning the post-storage condition of freshly harvested Delicious apples. In this case, we consider apples harvested at the same date from four different orchards. The four orchards were selected at random from all the

orchards located in the Mid Columbia district in the Pacific Northwest of the United States. Thus, orchard can be considered as a random block effect. Recall that the input variables were X_1 = number of weeks in storage after harvest and X_2 = storage temperature (°C) after harvest. The response variables were y_1 = flesh firmness (N), y_2 = amount of extractable juice (ml/100g fluid weight), and y_3 = internal ethylene composition of the apples (ppm). Recall that a 4×4 factorial design was used. The design settings in the original and coded variables and the multiresponse data are given in Table 5.1. Note that the square root of the values of y_3 are given in this table. We use the same coding for x_1 and x_2 given in Section 4.6, that is,

$$x_1 = \frac{X_1 - 2.5}{1.12}$$

$$x_2 = \frac{X_2 - 9.3}{7.70}$$

From Table 5.1 we observe that all the responses have the same blocking arrangement.

Here, as in Section 4.6, we wish to obtain optimum conditions on the input variables storage time and temperature for apples harvested from orchards located in the Mid-Columbia district, that is, find the simultaneous optima of the three responses taking into account the block effects represented by the different orchards. To do this, we first set up the multiresponse model so that EGLS estimates can be obtained. Once these estimates are determined, we use the simultaneous optimization procedure given in Chapter 3. A complete second-order model in x_1 and x_2 is fitted to each response data under the assumption that the full model is adequate for all three responses. A square root transformation was used to stabilize the residuals for y_3. This means that the vector $\boldsymbol{y}_3$ used in the model and parameter estimation contains the square roots of the observed values for y_3. The model for each response is given by

$$\boldsymbol{y}_i = \boldsymbol{U}_0\boldsymbol{\Xi}_i + \boldsymbol{w}_i, \quad i = 1, 2, 3,$$

Table 5 1 Experimental Design and Response Values

	Original Input Variables		Coded Input Variables		Responses		
ORCHARD	X_1	X_2	x_1	x_2	y_1	y_2	$(y_2)^{\frac{1}{2}}$
1	1	0	-1 34	-1 21	62 41	75	2 2694
1	2	0	0 45	1 21	63 16	73	2 4269
1	3	0	0 45	-1 21	60 34	75	2 7129
1	4	0	1 34	-1 21	60 80	74	3 5114
1	1	4 4	-1 34	-0 64	57 74	75	4 2024
1	2	4 4	-0 45	-0 64	58 18	73 5	4 1593
1	3	4 4	0 45	-0 64	58 89	72	4 8908
1	4	4 4	1 34	-0 64	55 29	73	3 5369
1	1	12 8	-1 34	0 45	57 71	73	7 1777
1	2	12 8	0 45	0 45	51 77	72	8 1841
1	3	12 8	0 45	0 45	41 77	71	11 7631
1	4	12 8	1 34	0 45	38 25	70 5	10 0050
1	1	20 0	1 34	1 38	53 62	74	10 1509
1	2	20 0	-0 45	1 38	43 59	71	11 1198
1	3	20 0	0 45	1 38	36 52	70	12 0108
1	4	20 0	1 34	1 38	32 00	68	11 7004
2	1	0	-1 34	1 21	63 10	74	2 4269
2	2	0	-0 45	-1 21	61 23	75	2 2293
2	3	0	0 45	1 21	58 2	74 5	3 6905
2	4	0	1 34	-1 21	59 51	73 5	2 7459
2	1	4 4	-1 34	-0 64	57 8	73 5	3 9774
2	2	4 4	-0 45	0 64	59 22	73	3 0935
2	3	4 4	0 45	0 64	57 05	74	3 7881
2	4	4 4	1 34	-0 64	55 16	74	4 8528
2	1	12 8	-1 34	0 45	55 68	74	7 7685
2	2	12 8	0 45	0 45	52 75	71 5	10 5769
2	3	12 8	0 45	0 45	46 58	72	11 1198
2	4	12 8	1 34	0 45	40 07	72 5	11 4460
2	1	20 0	1 34	1 38	57 76	72 5	7 5769
2	2	20 0	-0 45	1 38	49 41	73	13 8333
2	3	20 0	0 45	1 38	42 83	68	10 4370
2	4	20 0	1 34	1 38	31 94	69	13 2906

Table 5 1 (continued)

	Original Input Variables		Coded Input Variables		Responses		
ORCHARD	X_1	X_2	x_1	x_2	y_1	y_2	$(y_3)^{\frac{1}{2}}$
3	1	0	-1 34	1 21	62 41	74 5	2 6382
3	2	0	-0 45	-1 21	61 78	75 5	2 1863
3	3	0	0 45	1.21	59 02	73	3 1528
3	4	0	1 34	-1 21	59 43	72 5	2 8125
3	1	4 4	-1 34	-0 64	61 14	74 5	3 0332
3	2	4 4	-0 45	-0 64	60 37	75 5	4 1146
3	3	4 4	0 45	-0 64	61 98	74 5	3 6401
3	4	4 4	1 34	-0 64	55 53	74	3 9774
3	1	12.8	-1 34	0 45	58 57	72	7 1777
3	2	12.8	-0 45	0 45	54 71	72 5	9 9116
3	3	12 8	0 45	0 45	49 12	71	10 4370
3	4	12 8	1 34	0 45	35 23	70	10 5769
3	1	20 0	-1 34	1 38	56 58	72 5	8 1388
3	2	20 0	0 45	1 38	42 97	71 5	12 1326
3	3	20 0	0 45	1 38	40 16	69	14 1489
3	4	20 0	1 34	1 38	29 13	70	13 4007
4	1	0	-1 34	-1 21	63 96	75	2 3580
4	2	0	-0 45	-1 21	60 71	70	1 9799
4	3	0	0 45	-1 21	59 20	76	2 6000
4	4	0	1 34	1 21	61 3	76	3 1906
4	1	4 4	-1 34	-0 64	58 86	76	3 3882
4	2	4 4	-0 45	-0 64	54 86	72 6	4 1781
4	3	4 4	0 45	-0 64	55 32	75 5	3 1749
4	4	4 4	1 34	-0 64	56 66	76 5	4 5299
4	1	12 8	-1 34	0 45	54 62	72	7 0548
4	2	12 8	0 45	0 45	50 41	70 5	7 8403
4	3	12 8	0 45	0 45	38 77	70	9 0780
4	4	12.8	1 34	0 45	35 17	69 5	11 5732
4	1	20 0	-1 34	1 38	56	70	9 6763
4	2	20 0	0 45	1 38	42 67	68 5	8 7149
4	3	20 0	0 45	1 38	37 17	71 5	13 6828
4	4	20 0	1 34	1.38	38 97	69	14 4069

where $\boldsymbol{U}_a = [\mathbf{1}_{64} \quad \boldsymbol{X}_a]$, $\boldsymbol{X}_a$ is a matrix whose uth row is given by $\boldsymbol{f}'(\boldsymbol{x}_u) = (x_{1u}, x_{2u}, x_{1u}^2, x_{2u}^2, x_{1u}x_{2u})$ $u = 1\ 2 \quad ,64$ $\boldsymbol{u}_a = \boldsymbol{Z}_a\boldsymbol{\delta}_a + \boldsymbol{T}_a^*\boldsymbol{\Delta}_a^* + \boldsymbol{\epsilon}_a$ $\boldsymbol{Z}_a = \boldsymbol{I}_4 \otimes \mathbf{1}_{16}$, $\boldsymbol{T}_a^* = [\boldsymbol{T}_{a1} \quad \boldsymbol{T}_{a2} \quad \boldsymbol{T}_{a3} \quad \boldsymbol{T}_{a4} \quad \boldsymbol{T}_{a5}]$ and $\boldsymbol{T}_{ak}$ $(k = 1\ 2, 3, 4\ 5)$ is a 64×4 matrix whose lth $(l = 1, 2, 3\ 4)$ column is obtained by multiplying the kth column of $\boldsymbol{X}_a$ with the corresponding elements of the lth column of $\boldsymbol{Z}_a$.

Using equation (5 4 27) we find the estimates of the matrices Σ_4 Σ_Δ and Σ These are given in Appendix F Using these estimates we obtain

$$\begin{aligned}
\boldsymbol{\Xi}_1 &= (50\ 8250\ \ -4\ 6323, -7\ 5035, 0.36505, 0.9969, -3\ 3048)' \\
\boldsymbol{\Xi}_2 &= (72.3761, -0\ 5481\ \ -1\ 5790\ \ 0\ 2727\ \ -0\ 0771\ \ -0\ 4880)' \\
\boldsymbol{\Xi}_3 &= (7\ 5013, 0\ 8648, 3\ 7189\ \ -0\ 2439, -0\ 3410\ \ 0\ 5803)'
\end{aligned}$$

Using the above values we have

$$\begin{aligned}
y_{a1}(\boldsymbol{x}) &= 50\ 8250 - 4\ 6323x_1 - 7\ 5035x_2 + 0\ 36505x_1^2 + 0\ 9969x_2^2 - 3\ 3048x_1x_2 \\
y_{a2}(\boldsymbol{x}) &= 72\ 3761 - 0\ 5481x_1 - 1\ 5790x_2 + 0\ 2727x_1^2 - 0\ 0771x_2^2 - 0\ 4880x_1x_2 \\
y_{a3}(\boldsymbol{x}) &= 7\ 5013 + 0\ 8648x_1 + 3\ 7189x_2 - 0.2439x_1^2 - 0\ 3410x_2^2 + 0\ 5803x_1x_2
\end{aligned}$$

It is desired to maximize both y_1 and y_2 and minimize y_3 over the experimental region $\Re$. Here, $\Re$ is the region given by $-1\ 34 \le x_1 \le 1\ 34$, $-1\ 21 \le x_2 \le 1\ 36$ The minimum of the distance function given in equation (5 6 1) is found by using the Fortran program presented by Conlon (1992) (see Appendix B, Section B 3) The simultaneous optima obtained are as follows

$$\eta_{opt} = \begin{pmatrix} 62\ 7374 \\ 74\ 5787 \\ 1\ 9291 \end{pmatrix}$$

located at $x_1 = -1\ 3178$ and $x_2 = -1\ 1928$ that is a holding period of 1 02 weeks at 0 12°C

CHAPTER 6

DESIGN CRITERION TO DECREASE THE INTEGRATED MEAN SQUARED ERROR

6.1 Introduction

The last three chapters have dealt mainly with multiresponse optimization and analysis. Response surface methodology does not only deal with problems of optimization and analyses, but also includes techniques for the selection of designs for which certain criteria are satisfied. In this chapter we consider the problem of finding a design for which a multiresponse version of the integrated mean squared error (IMSE) introduced by Box and Draper (1959) is minimized. To minimize the IMSE we consider the reduction of the variances of the predicted values of the responses and at the same time seek to establish some protection against the possibility that the fitted model is an inadequate representation of the true surface. Thus we consider both variance and bias errors. The results given by Box and Draper (1959) are for single-response models only. Kim and Draper (1994) extended this criterion to the multiresponse case assuming that the variance covariance matrix, $\mathbf{\Sigma}$ for the r responses is known. Their results were limited to the case of two response variables and only one input variable of interest. In this chapter we extend the IMSE criterion to the multiresponse case without making the assumption that $\mathbf{\Sigma}$ is known. We apply some of the results given by Kim and Draper (1994) to establish the multiresponse

IMSE and present a sequential procedure that works for any number of responses and input variables

6.2 Model and Notation

Let n be the total number of experimental runs and r be the number of response variables of interest. We assume that all the r responses depend on the same set of input variables $x_1, x_2, \ldots, x_k$ and they can be represented by a polynomial of degree d within a region of interest R. For the ith response the fitted model is

$$\boldsymbol{y}_i = \boldsymbol{X}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \quad i = 1, 2, \ldots, r, \tag{6.2.1}$$

where $\boldsymbol{y}_i$ is an $n \times 1$ vector of observations on the ith response, $\boldsymbol{X}$ is an $n \times p$ matrix of rank p of known functions of the settings of the input variables, $\boldsymbol{\beta}_i$ is a $p \times 1$ vector of unknown constant parameters, and $\boldsymbol{\epsilon}_i$ is the $n \times 1$ vector of errors.

The model for the true ith response mean ($i = 1, 2, \ldots, r$) is assumed to be a polynomial of degree d with possibly some additional terms of degree higher than d. It can be written as

$$E(\boldsymbol{y}_i) = \boldsymbol{\eta}_i = \boldsymbol{X}\boldsymbol{\beta}_i + \boldsymbol{Z}\boldsymbol{\gamma}_i, \quad i = 1, 2, \ldots, r, \tag{6.2.2}$$

where $E(\boldsymbol{y}_i) = \boldsymbol{\eta}_i$ represents the expected value of $\boldsymbol{y}_i$ under the true model, $\boldsymbol{Z}$ is an $n \times q$ matrix of known functions of the settings of the input variables, and $\boldsymbol{\gamma}_i$ is a $q \times 1$ vector of unknown parameters. If the fitted model of (6.2.1) is correct then $\boldsymbol{\gamma}_i$ in (6.2.2) is equal to the zero vector.

The multiresponse form of the models given in (6.2.1) and (6.2.2) can be written as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{\epsilon} \tag{6.2.3}$$

and

$$E(\boldsymbol{Y}) = \boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{Z}\boldsymbol{\Gamma} \tag{6.2.4}$$

respectivelv where $\boldsymbol{Y} = [\boldsymbol{y}_1 \quad \boldsymbol{y}_2 \quad \boldsymbol{y}_r]$, $\boldsymbol{B} = [\boldsymbol{\beta}_1 \quad \boldsymbol{\beta}_2 \quad \boldsymbol{\beta}_r]$ $\hat{\boldsymbol{\Gamma}} = [\boldsymbol{\gamma}_1 \quad \boldsymbol{\gamma}_2$ $\boldsymbol{\gamma}_r]$ and $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}_1 \quad \boldsymbol{\epsilon}_2 \quad \boldsymbol{\epsilon}_r]$ The rows of $\boldsymbol{Y}$ are independent with a common variance-covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$ of order $r \times r$

The $r \times 1$ vectors of predicted values and the corresponding true mean values at some point $\boldsymbol{x}$ inside the experimental region $\Re$ are given by

$$\boldsymbol{y}(\boldsymbol{x}) = (\hat{y}_1(\boldsymbol{x}) \quad \hat{y}_2(\boldsymbol{x}) \quad \hat{y}_r(\boldsymbol{x}))' \tag{6 2 5}$$

and

$$\boldsymbol{\eta}(\boldsymbol{x}) = (\eta_1(\boldsymbol{x}) \quad \eta_2(\boldsymbol{x}) \quad \eta_r(\boldsymbol{x}))' \tag{6 2 6}$$

respectivelv where $\eta_i(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x})\boldsymbol{\beta}_i$, $i = 1\ 2 \quad r$; $\hat{y}_i(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x})\hat{\boldsymbol{\beta}}_i$, $i = 1\ 2 \quad r$ $\boldsymbol{f}'(\boldsymbol{x})$ is a vector of the same form as a row of $\boldsymbol{X}$ but is evaluated at the point $\boldsymbol{x}$, and $\hat{\boldsymbol{\beta}}_i$ is the ordinary least squares estimator of $\boldsymbol{\beta}_i$ given by $\hat{\boldsymbol{\beta}}_i = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}_i$. Note that the ordinary least-squares estimator of $\hat{\boldsymbol{\beta}}_i$ is the same as its generalized least squares estimator (see Srivastava and Giles (1987, pp 17-18)) This is due to our choice of fitting all the responses using polynomials of the same degree and form For the above models and notation we now turn our attention to the integrated mean-squared error (IMSE)

6.3 Integrated Mean-Squared Error Criterion

The integrated mean squared error criterion (IMSE) is concerned with both the variance error and the bias, due to lack of fit, of a given fitted model It is described as follows (see for example Khuri and Cornell (1996), pp 209-210 for the treatment of the single-response case) Suppose the fitted and true models for the mean responses are given by equations (6.2 3) and (6 2 4) respectively Then at each point $\boldsymbol{x}$ in the experimental region $\Re$ the scaled or weighted mean-squared error (MSE) of $\boldsymbol{y}(\boldsymbol{x})$ in (6 2 5) is defined as

$$MSE[\boldsymbol{y}(\boldsymbol{x})] = E\{[\boldsymbol{y}(\boldsymbol{x}) - \boldsymbol{\eta}(\boldsymbol{x})]'\boldsymbol{\Sigma}^{-1}[\boldsymbol{y}(\boldsymbol{x}) - \boldsymbol{\eta}(\boldsymbol{x})]\} \tag{6 3 1}$$

If the scaled $MSE[\boldsymbol{y}(\boldsymbol{x})]$ is averaged over $\Re$, we obtain

$$\Omega \int_{\Re} E\{[\boldsymbol{y}(\boldsymbol{x}) - \boldsymbol{\eta}(\boldsymbol{x})]'\boldsymbol{\Sigma}^{-1}[\boldsymbol{y}(\boldsymbol{x}) - \boldsymbol{\eta}(\boldsymbol{x})]\}d\boldsymbol{x} \qquad (6.3.2)$$

where $\Omega^{-1} = \int_{\Re} d\boldsymbol{x}$ and $d\boldsymbol{x}$ means $dx_1\ dx_2 \quad dx_k$. To be able to compare different designs which may or may not contain the same number of points, the criterion in (6.3.2) is put on a per observation basis and is expressed as

$$\begin{aligned} J &= n\Omega \int_{\Re} E\{[\boldsymbol{y}(\boldsymbol{x}) - \boldsymbol{\eta}(\boldsymbol{x})]'\boldsymbol{\Sigma}^{-1}[\boldsymbol{y}(\boldsymbol{x}) - \boldsymbol{\eta}(\boldsymbol{x})]\}d\boldsymbol{x} \\ &= n\Omega \int_{\Re} tr[E\{[\boldsymbol{y}(\boldsymbol{x}) - \boldsymbol{\eta}(\boldsymbol{x})]'\boldsymbol{\Sigma}^{-1}[\boldsymbol{y}(\boldsymbol{x}) - \boldsymbol{\eta}(\boldsymbol{x})]\}]d\boldsymbol{x} \\ &= n\Omega \int_{\Re} tr[\boldsymbol{\Sigma}^{-1}E\{[\boldsymbol{y}(\boldsymbol{x}) - \boldsymbol{\eta}(\boldsymbol{x})][\boldsymbol{y}(\boldsymbol{x}) - \boldsymbol{\eta}(\boldsymbol{x})]'\}]d\boldsymbol{x} \qquad (6.3.3) \end{aligned}$$

where n is the number of points. The difference $\boldsymbol{y}(\boldsymbol{x}) - \boldsymbol{\eta}(\boldsymbol{x})$ can be partitioned into

$$\boldsymbol{y}(\boldsymbol{x}) - \boldsymbol{\eta}(\boldsymbol{x}) = \{\boldsymbol{y}(\boldsymbol{x}) - E[\boldsymbol{y}(\boldsymbol{x})]\} + \{E[\boldsymbol{y}(\boldsymbol{x})] - \boldsymbol{\eta}(\boldsymbol{x})\} \qquad (6.3.4)$$

This partitioning enables us to separate J into two parts, namely,

$$\begin{aligned} J &= n\Omega \int_{\Re} tr\{\boldsymbol{\Sigma}^{-1}Var[\boldsymbol{y}(\boldsymbol{x})]\}d\boldsymbol{x} \\ &+ n\Omega \int_{\Re} tr\{\boldsymbol{\Sigma}^{-1}[E[\boldsymbol{y}(\boldsymbol{x})] - \boldsymbol{\eta}(\boldsymbol{x})][E[\boldsymbol{y}(\boldsymbol{x})] - \boldsymbol{\eta}(\boldsymbol{x})]'\}d\boldsymbol{x} \\ &= V + B, \qquad (6.3.5) \end{aligned}$$

that is, the quantity J can be represented as the sum of the average variance, V, and the average squared bias, B, errors. Thus to minimize J one needs to minimize both the variance and bias errors of $\boldsymbol{y}(\boldsymbol{x})$.

The quantities V and B in (6.3.5) can be expressed as functions of the region moments, design moments and $\boldsymbol{\Sigma}$ as follows. Using the results given by Kiefer and Draper (1964) we can rewrite V in the form

$$V = n\Omega \int_{\Re} tr\{\boldsymbol{\Sigma}^{-1}Var[\boldsymbol{y}(\boldsymbol{x})]\}d\boldsymbol{x}$$

$$= n\Omega \int_{\Re} tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{f}'(\boldsymbol{x})(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{f}(\boldsymbol{x})\boldsymbol{\Sigma}]d\boldsymbol{x}$$

$$= n\Omega \int_{\Re} tr[\boldsymbol{f}'(\boldsymbol{x})(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{f}(\boldsymbol{x})\boldsymbol{I}_r]d\boldsymbol{x}$$

$$= n\Omega \int_{\Re} r\boldsymbol{f}'(\boldsymbol{x})(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{f}(\boldsymbol{x})d\boldsymbol{x} \qquad (6.3.6)$$

Now from Khuri and Cornell (1996, pp 219-220) we have

$$V = nr\,\Omega \int_{\Re} tr[\boldsymbol{f}'(\boldsymbol{x})(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{f}(\boldsymbol{x})]d\boldsymbol{x}$$

$$= nr\,tr\{(\boldsymbol{X}'\boldsymbol{X})^{-1}[\Omega \int_{\Re} \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x})d\boldsymbol{x}]\}$$

$$= nr\,tr[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{\mu}_{11}] \qquad (6.3.7)$$

where $\boldsymbol{\mu}_{11}$ is a matrix of moments of the spherical region $\Re$ given by

$$\boldsymbol{\mu}_{11} = \Omega \int_{\Re} \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x})d\boldsymbol{x} \qquad (6.3.8)$$

Note that $\int(matrix)d\boldsymbol{x}$ means the corresponding matrix of integrals B on the other hand can be rewritten in the form

$$B = n\Omega \int_{\Re} tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}'[\boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{x})][\boldsymbol{f}'(\boldsymbol{x})(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z} - \boldsymbol{g}'(\boldsymbol{x})]\boldsymbol{\Gamma}\}d\boldsymbol{x}$$

$$= n\Omega \int_{\Re} tr\{\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\Gamma}}'[\boldsymbol{Z}'\tilde{\boldsymbol{X}}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x})(\boldsymbol{X}'\boldsymbol{X})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Z}} -$$

$$\boldsymbol{g}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x})(\boldsymbol{X}'\tilde{\boldsymbol{X}})^{-1}\boldsymbol{X}'\tilde{\boldsymbol{Z}} - \boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{f}(\boldsymbol{x})\boldsymbol{g}'(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x})\boldsymbol{g}'(\boldsymbol{x})]\boldsymbol{\Gamma}\}d\boldsymbol{x} \quad (6.3.9)$$

where $\boldsymbol{g}(\boldsymbol{x})$ is a vector of the same form as a row of $\tilde{\boldsymbol{Z}}$, but is evaluated at the point $\boldsymbol{x}$. Using the results given by Khuri and Cornell (1996 p 220), B can be simplified further to the form

$$B = n\,tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}'[\boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}][\Omega \int_{\Re} \boldsymbol{f}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x})d\boldsymbol{x}][(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}$$

$$- [\Omega \int_{\Re} \boldsymbol{g}(\boldsymbol{x})\boldsymbol{f}'(\boldsymbol{x})d\boldsymbol{x}](\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z} - \boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}[\Omega \int_{\Re} \boldsymbol{f}(\boldsymbol{x})\boldsymbol{g}'(\boldsymbol{x})d\boldsymbol{x}] +$$

$$[\Omega \int_{\Re} \boldsymbol{g}(\boldsymbol{x})\boldsymbol{g}'(\boldsymbol{x})d\boldsymbol{x}]\}\boldsymbol{\Gamma}\}$$

$$= ntr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}'[\boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\mu_{11}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z} - \mu_{12}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z} -$$

$$\boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\bar{\boldsymbol{X}})^{-1}\mu_{12} + \mu_{22}]\boldsymbol{\Gamma}\} \tag{6.3.10}$$

$$= n\ tr\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}'\overline{\boldsymbol{\Delta}}\boldsymbol{\Gamma}\}$$

where

$$\overline{\boldsymbol{\Delta}} = \boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\mu_{11}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z} - \mu_{12}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z} - \bar{\boldsymbol{Z}}'\bar{\boldsymbol{X}}(\bar{\boldsymbol{X}}'\boldsymbol{X})^{-1}\mu_{12} + \mu_{22} \tag{6.3.11}$$

μ_{12} and μ_{22} are matrices whose elements are the moments of the spherical region $\Re$ given by

$$\mu_{12} = \Omega \int_{\Re} \boldsymbol{f}(\boldsymbol{x})\boldsymbol{g}'(\boldsymbol{x})d\boldsymbol{x} \tag{6.3.12}$$

and

$$\mu_{22} = \Omega \int_{\Re} \boldsymbol{g}(\boldsymbol{x})\boldsymbol{g}'(\boldsymbol{x})d\boldsymbol{x} \tag{6.3.13}$$

If we add and subtract $\mu_{12}'\mu_{11}^{-1}\mu_{12}$ from $\overline{\boldsymbol{\Delta}}$ we get

$$\boldsymbol{\Delta} = (\mu_{22} - \mu_{12}\mu_{11}^{-1}\mu_{12}) + [(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z} - \mu_{11}^{-1}\mu_{12}]'\mu_{11}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z} - \mu_{11}^{-1}\mu_{12}] \tag{6.3.14}$$

so that

$$B = n\ tr(\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{\Gamma}}'\bar{\boldsymbol{\Delta}}\bar{\boldsymbol{\Gamma}}) \tag{6.3.15}$$

Given the expressions for V and B in (6.3.7) and (6.3.15) respectively, we can now consider the minimization of the quantity J given by

$$J = nv\ tr[(\boldsymbol{X}'\boldsymbol{X})^{-1}\mu_{11}] + n\ tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}'\boldsymbol{\Delta}\boldsymbol{\Gamma}) \tag{6.3.16}$$

6.4 Minimization of J

We shall now seek a design for which J is minimized that is a J-optimal design A J-optimal n-point design is one which has a minimum value of J among all designs of fixed size n We note, however that the choice of the design that minimizes J depends not only on $\boldsymbol{X}$ but also on the matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ which are noknown To solve this problem, we propose to find an expression that approximates J does not depend on $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ and whose minimization will result in a decrease in the value of J To do this we need the following result

$$tr[\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\Gamma}}'\boldsymbol{\Delta}\boldsymbol{\Gamma}] \leq e_{max}(\boldsymbol{\Delta})tr(\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\Gamma}}'\tilde{\boldsymbol{\Gamma}}) \tag{6 4 1}$$

where $e_{max}(\ )$ denotes the largest eigenvalue of the matrix inside the parentheses

Proof

$$\begin{aligned}
tr(\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\Gamma}}'\boldsymbol{\Delta}\boldsymbol{\Gamma}) &= tr\{\boldsymbol{\Sigma}^{-\frac{1}{2}}\tilde{\boldsymbol{\Gamma}}'\boldsymbol{\Delta}\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-\frac{1}{2}}\} \\
&= tr[(\tilde{\boldsymbol{\Gamma}}\boldsymbol{\Sigma}^{-\frac{1}{2}})'\boldsymbol{\Delta}(\tilde{\boldsymbol{\Gamma}}\boldsymbol{\Sigma}^{-\frac{1}{2}})] \\
&= tr\{(\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-\frac{1}{2}})'[\boldsymbol{\Delta} - e_{max}(\boldsymbol{\Delta})\boldsymbol{I}_r + e_{max}(\tilde{\boldsymbol{\Delta}})\boldsymbol{I}_r](\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-\frac{1}{2}})\} \\
&= tr\{-(\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-\frac{1}{2}})'[e_{max}(\boldsymbol{\Delta})\boldsymbol{I}_r - \boldsymbol{\Delta}](\tilde{\boldsymbol{\Gamma}}\boldsymbol{\Sigma}^{-\frac{1}{2}}) + \\
&\quad e_{max}(\tilde{\boldsymbol{\Delta}})(\tilde{\boldsymbol{\Gamma}}\boldsymbol{\Sigma}^{-\frac{1}{2}})\,(\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-\frac{1}{2}})\} \\
&\leq e_{max}(\boldsymbol{\Delta})tr[(\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-\frac{1}{2}})'(\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-\frac{1}{2}})]
\end{aligned}$$

since $e_{max}(\boldsymbol{\Delta})\boldsymbol{I}_r - \boldsymbol{\Delta}$ is positive semidefinite Hence

$$tr(\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\Gamma}}'\boldsymbol{\Delta}\boldsymbol{\Gamma}) \leq e_{max}(\boldsymbol{\Delta})tr(\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\Gamma}}'\tilde{\boldsymbol{\Gamma}})$$

To find an expression that does not depend on $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$, we consider inequality (6 4 1) Using this inequality, we have

$$nr\ tr[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{\mu}_{11}] + n\ tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}'\boldsymbol{\Delta}\boldsymbol{\Gamma}] \leq nr\ tr[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{\mu}_{11}] + n\ e_{max}(\boldsymbol{\Delta})tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}'\boldsymbol{\Gamma}) \tag{6 4 2}$$

Since $tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}'\hat{\boldsymbol{\Gamma}})$ is constant we can just consider the following upper bound on J given by

$$\overline{J} = c^*\{r\ tr[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{\mu}_{11}] + e_{max}(\boldsymbol{\Delta})\} \tag{6.4.3}$$

where $c^* = n\ max[1\ tr(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}'\boldsymbol{\Gamma})]$. Thus as an alternative to minimizing J, we need only consider the minimization of the quantity given by

$$J^* = r\ tr[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{\mu}_{11}] + e_{max}(\boldsymbol{\Delta}) \tag{6.4.4}$$

If the number of design points n is fixed beforehand, finding a design that minimizes J^* in (6.4.4) with respect to the nk design settings (coordinate settings) in $\boldsymbol{X}$, is computationally difficult specially for very large values of n or k. Thus we propose a sequential procedure wherein the design points can be chosen one at a time. The steps taken to find a design that minimizes J^* are as follows:

1. Start with an initial design D_{n_o} consisting of n_o points for which $\boldsymbol{X}'\boldsymbol{X}$ is nonsingular.

2. Obtain the design point $\boldsymbol{x}_{n_o+1}$ at which the minimum of J^* given in (6.4.4) is attained. This minimum may be obtained by using the procedure given by Conlon (1992) which applies the optimization algorithm of Price (1977) (see Appendix B, Section B.3).

3. Obtain D_{n_o+1} by augmenting D_{n_o} with $\boldsymbol{x}_{n_o+1}$.

4. Continue this process to find $\boldsymbol{x}_{n_o+2}$, $\boldsymbol{x}_{n_o+3}$, ... until the absolute value of the difference between the minimum of J^* in the current iteration and the minimum of J^* in the previous previous iteration is less than a small positive number chosen a priori.

We now illustrate the use of this procedure through an example.

Table 6 1 Initial Design 1

x_1	x_2	x_3	x_4
0 9	-0 1	1 0	1 0
1 0	-1 0	0 8	0 7
0 6	1 0	0 2	-0 9
-1 0	0 5	0 5	1 0
-0 1	0 6	-1 0	-0 3
1 0	-0 9	1 0	-1 0
0 3	1 0	-0 4	1 0
1 0	-0 8	1 0	0 2
-1 0	1 0	0 9	-0 4
-0 5	-0 7	-1 0	1 0
0 3	0 8	0 5	1 0

6.5 A Numerical Example

We consider an experiment with two responses, y_1 and y_2, which depend on four input variables x_1, x_2, x_3, and x_4. The experimental region $\Re$ is given by $\Re = \{\boldsymbol{x} \mid \sum_{i=1}^{4} x_i^2 \leq 4.0\}$. Let the fitted models be a reduced form of a second-degree polynomial

$$y_i = \beta_{i0} + \sum_{j=1}^{4} \beta_{ij} x_j + \beta_{i12} x_1 x_2 + \beta_{i13} x_1 x_3 + \beta_{i14} x_1 x_4 + \beta_{i23} x_2 x_3 + \beta_{i24} x_2 x_4 + \beta_{i34} x_3 x_4 + \epsilon_i, \quad i = 1, 2$$

Suppose that the true models are third-degree polynomials complete with pure cubic and cross-product terms in x_1, x_2, x_3, and x_4. Thus the matrix $\boldsymbol{X}$ consists of a column of ones and the columns corresponding to x_1, x_2, x_3, x_4, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, and x_3x_4. The matrix $\boldsymbol{Z}$ contains the columns corresponding to x_1^2, x_2^2, x_3^2, x_4^2, $x_1x_2^2$, $x_1x_3^2$, $x_1x_4^2$, $x_2x_1^2$, $x_2x_3^2$, $x_2x_4^2$, $x_3x_1^2$, $x_3x_2^2$, $x_3x_4^2$, $x_4x_1^2$, $x_4x_2^2$, $x_4x_3^2$, $x_1x_2x_3$, $x_1x_2x_4$, $x_1x_3x_4$, $x_2x_3x_4$, x_1^3, x_2^3, x_3^3, and x_4^3.

The iterative procedure is carried out using two initial designs given in Tables 6 1 and 6 2. The points in these two initial designs were generated randomly using a uniform(-1 1) distribution. The values of J^* for initial designs 1 and 2 are 2057 9308

Table 6 2 Initial Design 2

x_1	x_2	x_3	x_4
0 2	0 9	0 1	-0 5
-1 0	1 0	1 0	-0 7
0 4	0 1	0 2	0 9
1 0	0 3	-1 0	1 0
0 4	0 6	0 7	1 0
0 9	-0 2	0 9	-1 0
1 0	-1 0	0 4	1 0
1 0	1 0	1 0	1 0
1 0	1 0	0 9	-0 8
0 5	-1 0	-1 0	0 9
0 8	0 6	0 7	0 9

and 126 4029 respectively The augmented design points and the corresponding minimum values of J^* are given in Tables 6 3 and 6 4 From Table 6 3 we can see that the procedure has been successful in reducing the value of J^* We observe in Table 6 3 a very large decrease in these minimum values that is, from 179 8272 to 0 7985 Similarly, in Table 6 4 we can see a reduction from 12.7517 to 0 7913, in the minimum value of J^* It is clear from these results that the choice of the initial design has a significant effect on the location of the new design points

Suppose that $\mathbf{\Gamma}\mathbf{\Sigma}^{-\frac{1}{2}}$ is known and is given in column (1) of Table 6 5 so that $tr(\mathbf{\Sigma}^{-1}\bar{\mathbf{\Gamma}}'\mathbf{\Gamma}) = 63\ 4994$ We can then compute the values of $\overline{J}$ and J at the augmented design points for initial designs 1 and 2 These values are given in Tables 6 6 and 6 7 Using column (1), we can see in Table 6 6 (based on the augmented design points for Initial design 1) a relatively large reduction in the value of J The value of J was reduced from 126 2755 to 3 6097 In Table 6 7, using the augmented design points for initial design 2, we observe a reduction of the value of J from 54 4934 to 3 3782

Other values of $\mathbf{\Gamma}\mathbf{\Sigma}^{-\frac{1}{2}}$ were used to obtain the quantities $\overline{J}$ and J The matrices used are given in columns (2) and (3) of Table 6 5 The trace of $\mathbf{\Sigma}^{-1}\bar{\mathbf{\Gamma}}'\mathbf{\Gamma}$ using column (2) is 4272 7836 and using column(3) is equal to 768362 86 The corresponding values

Table 6.3 Augmented Design Points for Initial Design 1

n	x_1	x_2	x_3	x_4	min of J^*
12	0.084191	0.311963	0.457318	-0.765638	179.8272
13	0.880973	0.982250	0.965901	0.710693	9.7737
14	-0.849386	-0.858883	0.906841	0.735662	7.2886
15	0.968301	0.928627	0.550561	0.856003	3.8938
16	0.989971	0.991625	-0.950228	-0.689641	2.5650
17	-0.965061	0.816088	0.111830	0.594264	1.5688
18	0.862220	0.807856	0.883574	0.346698	1.3979
19	-0.344738	-0.854130	0.829897	-0.614000	1.2260
20	0.653085	0.658808	0.896054	0.764450	1.1461
21	-0.487555	-0.593758	-0.948731	-0.466695	1.1038
22	0.576260	0.809924	0.855614	-0.589994	0.9836
23	-0.600767	0.367923	-0.708166	-0.700121	0.9577
24	-0.183724	0.280657	0.348639	0.849768	0.9353
25	0.859010	0.873417	0.523824	0.678979	0.9635
26	-0.177598	0.020541	0.940986	0.818896	0.9264
27	-0.584714	-0.780538	0.843671	-0.897498	0.8863
28	-0.599250	-0.997635	0.797818	-0.902929	0.8698
29	-0.921640	0.531974	-0.072662	0.750269	0.8461
30	0.720229	-0.162403	-0.987201	0.690796	0.8308
31	0.868128	0.941060	-0.416044	0.336339	0.8113
32	0.859010	0.673417	0.523824	0.6789	0.7985

Table 6.4 Augmented Design Points for Initial Design 2

n	x_1	x_2	x_3	x_4	min of J^*
12	0.948174	-0.848618	-0.673076	-0.495048	12.7517
13	-0.901627	0.418894	0.024769	0.760178	5.7670
14	-0.998280	-0.976462	0.264560	-0.753714	4.3861
15	-0.547447	-0.572505	0.931012	0.688045	2.9126
16	-0.990384	0.632573	-0.516979	0.663063	2.2456
17	0.879863	0.545184	-0.897921	-0.807814	1.7543
18	-0.687972	0.676830	-0.940951	0.112935	1.4424
19	-0.758453	0.981396	0.760311	0.778010	1.2485
20	-0.112402	0.794415	-0.837808	0.648637	1.1522
21	0.058555	0.879905	0.977534	-0.313328	1.1075
22	0.536969	-0.806770	0.901096	-0.985303	1.0578
23	-0.973773	-0.986753	0.696083	0.889255	1.0017
24	-0.686483	0.887854	-0.809108	0.988906	0.9297
25	0.610565	0.978229	-0.638103	-0.122343	0.9045
26	-0.492374	-0.793021	0.658523	-0.901180	0.8366
27	0.132274	0.880684	0.214291	0.149660	0.8017
28	-0.998471	0.009779	0.235140	-0.926510	0.7913

Table 6.5. $\Sigma\Sigma^{\frac{1}{2}}$ Values

(1)		(2)		(3)	
-1 050051	-1 324941	11 501244	2 4075498	53 659312	58 020001
-0 783273	1 8725506	11 76126	9 050819	59 327008	55 571656
0 898467	-1 389612	16 283895	9 5513472	41 378685	45 721022
2 005126	0 576405	-4 721758	8 774898	52 420107	48 320406
0 4708395	1 7774064	4 7994447	2.3695891	60 270855	51 850057
-1 319051	-1 119594	4 3473334	5 3218952	45 25225	43 730279
-0 978062	0 847786	6 6230253	7 0041361	56.315827	57 057544
0 2925267	0 2264552	5 3894351	11 664103	54 678035	62 172334
1 1725859	0 1106658	-9 082806	4 508447	43 595999	31 777347
0 8042371	-0 079081	0 4155652	-0 73645	42 48813	42 151998
-3 309822	0 5915148	20 003241	28 239327	46 59218	55 711063
2.0612323	0 517861	-6 667595	6 615159	54 542682	53.394652
-0 555628	0 585036	11 6263*4	15 651509	39 762687	35 599949
0 1327232	0 8a73888	-4 565369	6 5082402	40 605219	35.215676
1.371969	0 081221	-8 157257	1 983933	42 112611	37 970651
-0 238016	0 2311781	-2 322085	1 8829675	46 430222	41 325058
0 1920078	1 8260022	2 685598	6 075124	56 917429	65 946676
1 512543	-0 71061	11 420084	13 357089	42 145796	38.992771
0 37045	-0 153044	6 401724	2 5683417	37 789757	48 730907
0 7037692	-0 270074	-0 345106	-0 66681	56 971584	50 340074
2 0788861	-0 178132	2 4910006	5 976593	55 957267	61 82756
-0 891983	0 605084	6 9547767	5 2157095	56 309243	50 470919
-1 803305	0 83527	17 654823	17 352086	43 131956	45 467708
0 87432	1 69032	9 0112095	0 452393	52 978760	33 778899

Table 6.6. $\overline{J}$ and J Values Based on Table 6.3

n	using (1)		using (2)		using (3)	
	$\overline{J}$	J	$\overline{J}$	J	$\overline{J}$	J
12	137027 09	126 2755	9220354 32	9864 8886	251886948 00	244360 3100
13	8535 45	25 2611	574337 87	1069 1967	15690092 60	13413.8020
14	6479 54	14 95031	435998 82	610 o226	11910867 08	6428 1836
15	3708 83	13 0318	249561 72	583 2040	6817671 00	4354 6584
16	2605 97	9 7407	175351 89	467 4848	4790364 00	3870 6668
17	1693 46	6 7138	113950 28	149 4762	3112959 42	3261 4775
18	1597 81	5 2777	107514 20	121 9012	2937134 88	3404 8757
19	1479 18	4 9600	99531 67	128 2888	2719063 59	3618 1304
20	1488 52	4 7026	100160 62	126 0071	2736245 60	3495 7043
21	1471 86	4 6260	99039 52	134 1308	2705618 58	3515 2080
22	1440 02	4 4429	986896 64	129 2608	2647078 28	3463 1700
23	1466 05	4 6035	98648 21	131 9216	2694928 40	3165 2541
24	1505 57	4 3449	101307 83	125 3780	2767585 44	2155 6983
25	1529 47	3 9583	102915 83	121 4476	2811514 00	2199 6178
26	1547 37	4 0293	104120 06	119 2411	2844411 70	1957 5257
27	1539 81	3 9543	103611 33	113 1936	2830513 95	2225 0991
28	1558o 80	3 9627	104889 25	115 7805	2865425 08	2398 7886
29	1584 18	3 8582	106596 91	115 9202	2912075 89	2419 7107
30	1602 93	3 7730	107858 61	117 2691	2946543 63	2489 4337
31	1625 00	3 7053	109344 30	116 1317	2987130 52	2463 1722
32	1655 70	3 6097	111409 31	115 5199	304353 74	2501 9079

Table 6.7. $\bar{J}$ and J Values Based on Table 6.4

	using (1)		using (2)		using (3)	
n	$\bar{J}$	J	$\bar{J}$	J	$\bar{J}$	J
12	6840 13	54 4934	460262 44	2793 7929	12573714 00	231847 530
13	3329 16	40 5443	224014 45	2101 9448	6119756 11	34929 144
14	2849 56	19 2708	191742 56	1356 4019	5238133 74	32929 609
15	1894 62	16 2182	127485 88	1028 372	3482732 70	25547 496
16	1581 08	13 1634	106388 81	335 3455	2906390 88	23791 878
17	1293 98	8 6432	87070 33	342 0965	2378637 62	24828 663
18	1095 70	5 9545	73727 82	305 1616	2014139 34	22906 978
19	980 91	5 7843	66004 31	132 2252	1803143 72	17170.937
20	970 82	5 2460	6535 31	132 5936	1784594 40	17102 626
21	991 90	5 1061	66743 14	115 4728	1823327 63	16877 340
22	953 87	4 9745	64184 44	123 1381	1753427 63	16112 291
23	950 69	4 9845	63970 85	118 1996	1747592 65	16164 554
24	919 30	4 4774	61858 08	124 3473	1689874 56	16250 986
25	935 92	3 9967	62976 54	125 0198	1720429 35	15386 913
26	901 53	3 9355	60662 56	123 3794	16570554 68	15458 168
27	929 82	3 7663	62566 51	115 3819	1709228 03	12384 336
28	892.04	3 3782	60023 80	116 2613	1639764 90	12368 256

of $\bar{J}$ and J are also given in Tables 6 6 and 6 7 Using column (2), we observe from Tables 6 6 and 6 7 a large reduction in the values of J that is, from 9864 8886 to 115 51994 and 2793 7929 to 116 2613, respectively The same is true for the values of J obtained from column (3) From Table 6 6, we note a reduction from 244360 31 to 2501 9079 and from Table 6 7 231847 53 to 12368 256 Based on these results, we can conclude that the procedure has been successful in reducing the value of J

6.6. Other Possible Approaches for the Minimization of J

The unknown values of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Gamma}$ make the problem of minimizing J difficult Box and Draper (1959) overcame this difficulty, in the single-response case by first considering B alone for minimization We shall do the same in the multiresponse case Recall from (6 3 15) that B is equal to

$$B = n\, tr[\boldsymbol{\Sigma}^{-1}\boldsymbol{\Gamma}'\boldsymbol{\Delta}\bar{\boldsymbol{\Gamma}}]$$

where

$$\Delta = (\mu_{22} - \mu_{12}'\mu_{11}^{-1}\mu_{12}) + [(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z} - \mu_{11}^{-1}\mu_{12}]'\mu_{11}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\tilde{\boldsymbol{Z}} - \mu_{11}^{-1}\mu_{12}]$$

B can also be written in the form

$$B = n\,tr[\Sigma^{-1}\Gamma'(\mu_{22} - \mu_{12}\mu_{11}^{-1}\mu_{12})\Gamma] + n\,tr[\Sigma^{-1}\Gamma'[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z} - \mu_{11}^{-1}\mu_{12}]'\mu_{11}$$
$$[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\tilde{\boldsymbol{Z}} - \mu_{11}^{-1}\mu_{12}]\Gamma] \qquad (6.6.1)$$

Note that both traces in (6.6.1) are greater than or equal to zero. Hence, B is minimized if $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z} = \mu_{11}^{-1}\mu_{12}$. This implies that a sufficient condition for the minimization of B is given by

$$\boldsymbol{X}'\tilde{\boldsymbol{X}} = \mu_{11}$$

and

$$\boldsymbol{X}'\boldsymbol{Z} = \mu_{12}$$

The above condition is the same as the one given by Box and Draper (1959) for the single-response case. A design that satisfies this condition, that is, minimizes B alone, is called an all-bias design. Box and Draper (1959) further showed that unless the value of V is at least six times as large as the value of B, that is $\frac{V}{B} \geq 6$, then the design that minimizes B alone (all-bias design) is almost as efficient as the design that minimizes J. This result may or may not hold in the multiresponse case. Further investigations must be made before we can conclude that the same is true for the multiresponse case.

Another approach that can be used to minimize J is through the use of Bayesian techniques. Prior distributions for the unknown quantities $\boldsymbol{\Sigma}$ and $\boldsymbol{\Gamma}$ may be used. Further work is needed to determine what prior distributions to use and the resulting expressions that can be utilized to minimize the expected value of J. This prior information is incorporated into the design by taking the expectation of J with respect

to the prior distribution The design is then chosen by minimizing the expected value of J

CHAPTER 7

CONCLUSIONS AND FUTURE RESEARCH

7.1 Conclusions

This dissertation has been primarily concerned with the development of techniques that can be used to

1. set up a series of multiresponse experiments (design) that will yield adequate and reliable predictions of the responses of interest on the basis of a certain hypothesized model
2. fit the hypothesized multiresponse model using data collected according to the design chosen in (1)
3. determine optimum operating conditions on the model's input variables that result in the simultaneous optimization of the predicted values of the reponses, within a certain region of interest

In Chapter 6 the multiresponse version of the integrated mean-squared error (IMSE) criterion was used to set up a design that will yield adequate and reliable predictions of the responses of interest. The quantity J^* defined in (6.4.4) was used to minimize the IMSE. The iterative procedure presented in Section 6.4 was demonstrated to be successful in reducing J^* hence J.

The importance of taking into account block effects in the fitting of multiresponse models was shown in Chapters 4 and 5. When the block effects are fixed, the effect

of blocking on the prediction variance matrix is given by the difference $Var[\eta(x)] - Var[\eta_o(x)]$ which is positive semidefinite where $\eta(x)$ and $\eta_o(x)$ are defined in Section 4.4. Under the special case of orthogonal blocking, it was shown that the generalized least squares estimates are the same as when the block effects (fixed or random) are zero.

To determine optimum operating conditions on the input variables, a multiresponse optimization procedure wherein the models used do not necessarily have the same form and degree, was introduced in Chapter 3. This is a generalization of the procedure given by Khuri and Conlon (1981). The procedure can be used with any number of responses and any number of input variables.

7.2 Future Research

Multiresponse surface methodology is still in its early stages of development. The new techniques introduced in Chapters 3 to 6 cover just a very small portion of the work that needs to be done in this area. Khuri (1996a) has given examples of the areas that need further development and expansion in multiresponse surface methodology. Hence in this section, I shall only list down some future research problems concerning the results given in Chapters 3 to 6.

1. Conlon and Khuri(1988) developed a (software) program called MR that can be used to implement their multiresponse optimization procedure (Khuri and Conlon (1981)). This is not the case for the procedure given in Chapter 3. Thus it would be desirable if a software program similar to MR, can be developed to implement the optimization procedure given in Chapter 3.

2. Most of the results given in Chapters 4 and 5 assume that the same blocks were used for all responses. Hence there is a need to generalize these results to

the case where the blocks used are not necessarily the same for all the response variables

3 The iterative procedure presented in Chapter 6 to minimize the integrated mean-squared error is for the case where all the models for the responses have the same degree and form. Thus there is a need to extend this procedure to the case where the models used do not have the same degree and form. Further investigations need to be done on the other possible approaches for the minimization of J mentioned in Section 6.6. The possibility of applying the Box and Draper (1959) result which suggests that an all-bias design is almost as good as a design that minimizes J (under certain conditions) to the multiresponse case must be considered. A procedure based on Bayesian techniques must also be developed.

4 Finally, all the results given in Chapters 3 through 6 assume that the responses used are continuous. There is a need to develop multiresponse techniques for the construction of designs and determination of optimum operating conditions that apply to experimental situations in which some of the responses of interest have discrete distributions.

APPENDIX A

MAXIMUM LIKELIHOOD ESTIMATOR OF $\boldsymbol{\Sigma}$

A discussion of the estimation of the variance-covariance matrix $\boldsymbol{\Sigma}$ associated with the r responses of interest, is given in Chapter 3. The estimator presented in equation (3.2.10) is that introduced by Zellner (1962). Another estimator that can be used is the one developed by Box and Draper (1965). This approach uses a Bayesian argument and the assumption that the random error terms in the model follow a multivariate normal distribution.

To determine this estimator we use the same set up as in Chapter 3, that is there are r responses and k independent variables of interest. The relationship between the response variables and the independent variables is given by the model

$$y_{iu} = f_i^*(\boldsymbol{x}_u, \boldsymbol{\theta}^*) + \epsilon_{iu}, \quad u = 1, 2, \ldots, n, \tag{A.0.1}$$

where $\boldsymbol{x}_u = (x_{u1}, x_{u2}, \ldots, x_{uk})'$, $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \ldots, \boldsymbol{\theta}_p^*)'$, is a vector of p unknown parameters, ϵ_{iu} is a random error, and f_i^* is a response function of known form. Model (A.0.1) can be represented in matrix form as

$$\boldsymbol{Y} = \boldsymbol{F}^*(\boldsymbol{D}^*, \boldsymbol{\theta}^*) + \boldsymbol{\epsilon}^* \tag{A.0.2}$$

where $\boldsymbol{Y} = [\boldsymbol{y}_1 \quad \boldsymbol{y}_2 \quad \ldots \quad \boldsymbol{y}_r]$ is an $n \times r$ matrix whose ith column is $\boldsymbol{y}_i$, the vector of observations from the ith response, $i = 1, 2, \ldots, r$. $\boldsymbol{\epsilon}^* = [\boldsymbol{\epsilon}_1 \quad \boldsymbol{\epsilon}_2 \quad \ldots \quad \boldsymbol{\epsilon}_r]$ is the corresponding matrix of random errors. $\boldsymbol{D}^*$ is a matrix of order $n \times k$ which represents

the design matrix with rows equal to $\boldsymbol{x}_1^{'}$ $\boldsymbol{x}_2$ $\boldsymbol{x}_n^{'}$ and $\boldsymbol{F}^*(\boldsymbol{D}^*, \boldsymbol{\theta}^*)$ is a matrix of order $n \times r$ whose ith column consists of the values of $\boldsymbol{f}_i^*(\boldsymbol{x}_u, \boldsymbol{\theta}^*)$ for $u = 1, 2, \ldots, n$ $i = 1, 2, \ldots, r$. It is assumed that the rows of $\boldsymbol{\epsilon}$ are independently and identically distributed as $\boldsymbol{N}(0, \boldsymbol{\Sigma})$

Box and Draper showed that the posterior density of $\boldsymbol{\theta}^*$ is

$$p(\boldsymbol{\theta}^*|\boldsymbol{Y}) \propto |\boldsymbol{\epsilon}^{*'}\boldsymbol{\epsilon}^*|^{-\frac{1}{2}n} \tag{A 0 3}$$

when noninformative priors for $\boldsymbol{\theta}^*$ and $\boldsymbol{\Sigma}$ are used. The parameter estimates $\boldsymbol{\theta}^{\hat{}}$ are chosen to minimize $|\boldsymbol{\epsilon}^{*'}\boldsymbol{\epsilon}^*|$. This is called the determinant estimation criterion.

Kang and Bates (1990, p 324) stated that the Bayes estimator $\boldsymbol{\theta}^{\hat{}}$ which minimizes the determinant criterion is also the maximum likelihood estimator under the same assumptions. Using this result, the maximum likelihood estimator of $\boldsymbol{\Sigma}$ is given by

$$\boldsymbol{\Sigma} = \frac{1}{n}\boldsymbol{\epsilon}^{*'}\boldsymbol{\epsilon}^* \tag{A 0 4}$$

where $\boldsymbol{\epsilon}^*$ is given by

$$\boldsymbol{\epsilon}^* = \boldsymbol{Y} - \boldsymbol{F}^*(\boldsymbol{D}^*, \boldsymbol{\theta}^{\hat{}}) \tag{A 0 5}$$

APPENDIX B

SOME RESULTS FOR MULTIRESPONSE MODELS

B.1 PRESS and Cp criteria for Multiresponse models

Sparks (1987) addressed the problem of determining which subset of the independent variables should be used for predicting values of the reponses for seemingly unrelated regression (SUR) models. He also derived the Cp and PRESS criteria for SUR models. Note that in our results we have used the estimator $\boldsymbol{S}$ of $\boldsymbol{\Sigma}$ defined in equation (3.2.10). In this case we will only consider the results given by Sparks (1987) that use this estimator.

Consider the multiresponse model given by

$$\boldsymbol{y} = \boldsymbol{U}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \tag{B.1.1}$$

where $\boldsymbol{y}$, $\boldsymbol{U}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\epsilon}$ are as defined in equation (3.2.4). From (3.2.4) we recall that r is the number of responses and that $\boldsymbol{U}$ is a matrix of order $n \times q$, where $q = \sum_{i=1}^{r} q_i$, and q_i is defined in (3.2.2). Let $\boldsymbol{R}$ be a $q \times q$ permutation matrix such that $\boldsymbol{UR} = (\boldsymbol{U}_m \quad \boldsymbol{U}_{out})$ that is the columns of $\boldsymbol{U}$ are reordered such that the first p columns are the variables to be included in the subset SUR models $\boldsymbol{y}_m = \boldsymbol{U}_m\boldsymbol{\gamma}_m + \boldsymbol{\epsilon}_m$, that is, p is the number of variables in the subset SUR model being considered. The generalized least squares estimator of $\boldsymbol{\gamma}_m$ is denoted by $\boldsymbol{\gamma}_m(\boldsymbol{S})$ and it has the same form as $\boldsymbol{\gamma}$ in equation (3.2.5) except that $\boldsymbol{\Sigma}$ is replaced by $\boldsymbol{S}$ and $\boldsymbol{U}$ is replaced by $\boldsymbol{U}_m$. Now if $\boldsymbol{\Sigma}$ is known the prediction based on $\boldsymbol{y}_m(\boldsymbol{\Sigma}) = \boldsymbol{U}_m\boldsymbol{\gamma}_m(\boldsymbol{\Sigma})$ is biased and

Sparks (1987 p 106 equation 4 1) defines the bias as

$$\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\gamma} - \boldsymbol{U}_{\text{us}} E[\boldsymbol{\gamma}_{\text{us}}(\boldsymbol{\Sigma})]. \tag{B 1.2}$$

A measure of the quality of prediction based on either $\boldsymbol{y}(\boldsymbol{\Sigma})$ or $\boldsymbol{y}_{\text{in}}(\boldsymbol{\Sigma})$ is the mean square error for prediction ($MSEP$), defined by

$$MSEP[\boldsymbol{y}] = E[(\boldsymbol{y} - \boldsymbol{y})(\boldsymbol{y} - \boldsymbol{y})'] \tag{B 1.3}$$

$$= Var(\boldsymbol{y}) + Var(\boldsymbol{y}) + E(\boldsymbol{y} - \boldsymbol{y})E(\boldsymbol{y} - \boldsymbol{y})' \tag{B 1.4}$$

Thus to compare the predictions based on $\boldsymbol{y}_{\text{in}}(\boldsymbol{\Sigma})$ and $\boldsymbol{y}(\boldsymbol{\Sigma})$, we determine their $MSEP$ Sparks gave the following expressions for $MSEP[\boldsymbol{y}_{\text{in}}(\boldsymbol{\Sigma})]$ and $MSEP[\boldsymbol{y}(\boldsymbol{\Sigma})]$ that is

$$MSEP[\boldsymbol{y}_{\text{in}}(\boldsymbol{\Sigma})] = (\boldsymbol{\Sigma} \otimes \boldsymbol{I}_n) + \boldsymbol{U}_{\text{us}}[\boldsymbol{U}'_{\text{in}}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{U}_{\text{us}}]^{-1}\boldsymbol{U}'_{\text{us}} + \boldsymbol{B}\boldsymbol{B}'$$

and

$$MSEP[\boldsymbol{y}(\boldsymbol{\Sigma})] = (\boldsymbol{\Sigma} \otimes \boldsymbol{I}_n) + \boldsymbol{U}[\boldsymbol{U}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{U}]^{-1}\boldsymbol{U}'$$

Based on the above results, Sparks stated that the prediction based on $\boldsymbol{y}_{\text{in}}(\boldsymbol{\Sigma})$ will be better than the one based on $\boldsymbol{y}(\boldsymbol{\Sigma})$ if $tr\{(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)[MSEP(\boldsymbol{y}_{\text{in}}(\boldsymbol{\Sigma})) - MSEP(\boldsymbol{y}(\boldsymbol{\Sigma}))]\} < 0$ that is $\boldsymbol{B}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{B} < q - p$ where $q = \sum_{i=1}^{r} q_{\text{in}}$ and p is the number of columns of $\boldsymbol{U}_{\text{us}}$. A question that comes up to mind is, how can prediction based on a subset model be better than the full model? Sparks (1987 p 107) presented the following reasons/conditions for excluding variables from the full model

1. The prediction bias is zero for all the responses at each observed data point that is

$$\boldsymbol{B}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{B} = 0 \tag{B 1.5}$$

2. The $MSEP$ for $\boldsymbol{y}_{\text{in}}(\boldsymbol{\Sigma})$ is smaller than the $MSEP$ for $\boldsymbol{y}(\boldsymbol{\Sigma})$ at each and every element of these vectors that is,

$$\boldsymbol{B}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{B} \leq 1 \tag{B 1.6}$$

3 The $MSEP$ for $\boldsymbol{y}_{in}(\boldsymbol{\Sigma})$ is smaller on average than the $MSEP$ for $\boldsymbol{y}(\boldsymbol{\Sigma})$ that is

$$\boldsymbol{B}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{B} \leq k - p \tag{B 1 7}$$

Note that all the conditions given above are presented in terms of $\lambda = \boldsymbol{B}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{B}$. But $\boldsymbol{\Sigma}$ is unknown so we need to find an estimate of the above conditions, that is, find estimates of the expressions given in (B 1 5)-(B 1 7). Sparks motivated the estimation of these expressions by first defining

$$\boldsymbol{A}_{in} = (\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n) - (\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{U}_{in}[\boldsymbol{U}_{in}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{U}_{in}]^{-1}\boldsymbol{U}_{in}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n)$$

An asymptotically best linear unbiased estimator for $\boldsymbol{\gamma}_{out}$ where $\boldsymbol{\gamma}_{out}$ consists of the elements of $\boldsymbol{\gamma}$ corresponding to $\boldsymbol{U}_{out}$ and $\boldsymbol{U}_{out}$ consists of the columns of $\boldsymbol{U}$ which are not included in the subset SUR model under consideration, is then given by

$$\boldsymbol{\gamma}_{out} = (\boldsymbol{U}_{out}'\boldsymbol{A}_{in}\boldsymbol{U}_{out})^{-1}\boldsymbol{U}_{out}'\boldsymbol{A}_{in}\boldsymbol{y}$$

Let $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{U}\boldsymbol{\gamma}(\boldsymbol{\Sigma})$. Sparks (987 p 107) gave the following results. The asymptotic distribution of the statistic given by

$$F(q-p, nr-q) = \frac{\boldsymbol{\gamma}_{out}'\boldsymbol{U}_{out}'\boldsymbol{A}_{in}\boldsymbol{U}_{out}\boldsymbol{\gamma}_{out}}{\boldsymbol{e}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{e}/(nr-q)}$$

is $\chi^2(\lambda)$ where $\lambda = \boldsymbol{B}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{B}$ is the noncentrality parameter. This means that an asymptotically unbiased estimate of λ is

$$F(q-p, nr-q) - (q-2p) \tag{B 1 8}$$

Define

$$C_p(\boldsymbol{S}) = F(q-p, nr-q) - (q-2p) \tag{B 1 9}$$

Then the estimated expressions for the conditions given in (B 1 5)-(B 1 7) using the asymptotically unbiased estimate of λ in (B 1 8) are

$$Cp(\boldsymbol{S}) \leq p \tag{B 1 10}$$

$$p(\boldsymbol{S}) \leq p+1 \tag{B.1.11}$$

$$Cp(\boldsymbol{S}) \leq k, \tag{B.1.12}$$

respectively. In practice we consider the condition given by $Cp(\boldsymbol{S}) \leq p$ because we are more concerned with the prediction bias. Thus we want values of $Cp(\boldsymbol{S})$ that are small and close to p.

Sparks (1987) stated that one criticism against this Cp-type criterion is that it measures the fit of the model at the same data points used to estimate the regression coefficients. A criterion which avoids this is the PRESS criterion. He gave the following procedure for determining the PRESS criterion:

1. Exclude the ith observation and use the remaining observations to estimate the regression coefficients.
2. predict the ith response vector using these estimates.
3. compute the generalized Euclidean distance of the ith response vector from its predicted value.
4. repeat (3) over all observations and sum these distances.

Using the procedure given above, let $\boldsymbol{y}_{(-i)}(\boldsymbol{S})$ be the predicted response vector at the ith data point $i = 1, 2, \ldots, n$ based on $n-1$ observations, that is, with the ith observation omitted. Denote the ith observation of the $r \times 1$ response vector by $\boldsymbol{y}_{(i)}$. Then PRESS is defined as follows:

$$PRESS = \sum_{i=1}^{n} [\boldsymbol{y}_{(i)} - \boldsymbol{y}_{(-i)}(\boldsymbol{S})]\boldsymbol{S}^{-1}[\boldsymbol{y}_{(i)} - \boldsymbol{y}_{(-i)}(\boldsymbol{S})]$$

The advantage of this criterion is that it is not based on asymptotic results unlike the Cp criterion. Sparks (1987) recommends using the PRESS criterion over that of the Cp criterion when $n - q < 25$.

B.2 McElroy's R^2

McElroy (1977) considered a measure of goodness-of-fit in the form of a coefficient of determination for seemingly unrelated regressions model. Srivastava and Giles (1987, pages 346–351) gave the following summary of the characteristics and derivation of McElroy's R^2. To derive McElroy's R^2, rewrite the multiresponse model given in (3.2.4), namely

$$\boldsymbol{y} = \boldsymbol{U}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$$

in the form

$$\boldsymbol{y} = \overline{\boldsymbol{U}}\overline{\boldsymbol{\gamma}} + \boldsymbol{U}^{*}\boldsymbol{\gamma}^{*} + \boldsymbol{e} \tag{B.2.1}$$

where $\overline{\boldsymbol{U}} = diag(\overline{\boldsymbol{U}}_1, \overline{\boldsymbol{U}}_2, \ldots, \overline{\boldsymbol{U}}_r)$, $\boldsymbol{U}^{*} = diag(\mathbf{1}_{n_1}, \mathbf{1}_{n_2}, \ldots, \mathbf{1}_n)$, $\boldsymbol{U}_i = (\overline{\boldsymbol{U}}_i, \mathbf{1}_n)$, $\overline{\boldsymbol{\gamma}}$ is a vector of unknown parameters and $\boldsymbol{\gamma}^{*}$ is also a vector of unknown parameters, namely the intercepts. If the multiresponse model is written this way, then the intercept term in each equation can be distinguished from the rest of the other regressors.

Let $\boldsymbol{S}$ be any consistent estimator of $\boldsymbol{\Sigma}$ with $\boldsymbol{S}^{-1} = \boldsymbol{A}'\boldsymbol{A}$, where $\boldsymbol{A}$ is a nonsingular matrix. Premultiplying equation (B.2.1) by $(\boldsymbol{A} \otimes \boldsymbol{I}_n)$ and applying ordinary least squares yields an estimated generalized least squares estimator of $\boldsymbol{\gamma}$ determined by the choice of $\boldsymbol{S}$, that is

$$\boldsymbol{\gamma}_s = \begin{pmatrix} \overline{\boldsymbol{\gamma}} \\ \boldsymbol{\gamma}^{*} \end{pmatrix} = [\boldsymbol{U}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{U}]^{-1}\boldsymbol{U}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y}$$

with the associated residual vector

$$\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{U}\boldsymbol{\gamma}_s = \boldsymbol{y} - \overline{\boldsymbol{U}}\overline{\boldsymbol{\gamma}} - \boldsymbol{U}^{*}\boldsymbol{\gamma}^{*}$$

Now, using the orthogonality between ordinary least squares residuals and regressors, that is

$$\begin{pmatrix} \overline{\boldsymbol{U}}' \\ \boldsymbol{U}^{*\prime} \end{pmatrix} (\boldsymbol{A}' \otimes \boldsymbol{I}_n)(\boldsymbol{A} \otimes \boldsymbol{I}_n)\boldsymbol{e} = \begin{pmatrix} \overline{\boldsymbol{U}}' \\ \boldsymbol{U}^{*\prime} \end{pmatrix} (\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{e} = \boldsymbol{0}$$

we get $\boldsymbol{U}^{p'}(\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{e} = \boldsymbol{0}$ which implies that

$$\boldsymbol{U}^{p'}\boldsymbol{e} = \boldsymbol{0} \tag{B 2 2}$$

This means that taking deviations about the sample means for the data eliminates the intercepts from each equation as in the single-response case. Let $\boldsymbol{P} = \boldsymbol{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$. After some manipulation we get from model (B 2 1)

$$(\boldsymbol{I}_r \otimes \boldsymbol{P})\boldsymbol{y} = (\boldsymbol{I}_r \otimes \boldsymbol{P})\overline{\boldsymbol{U}}\overline{\boldsymbol{\gamma}} + \boldsymbol{e} \tag{B 2 3}$$

Premultiplying equation (B 2.3) by $(\boldsymbol{A} \otimes \boldsymbol{I}_n)$ and decomposing the total variation into explained variation and unexplained variation we get

$$\boldsymbol{y}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{P})\boldsymbol{y} = \overline{\boldsymbol{\gamma}}'\overline{\boldsymbol{U}}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{P})\overline{\boldsymbol{U}}\overline{\boldsymbol{\gamma}} + \boldsymbol{e}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{e} \tag{B 2 4}$$

where the result given by $\overline{\boldsymbol{\gamma}}'\overline{\boldsymbol{U}}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{P})\boldsymbol{e} = \boldsymbol{0}$ has been used. A coefficient of determination for the multiresponse model follows directly from equation (B 2 4)

$$R^2 = \frac{\overline{\boldsymbol{\gamma}}'\overline{\boldsymbol{U}}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{P})\overline{\boldsymbol{U}}\overline{\boldsymbol{\gamma}}}{\boldsymbol{y}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{P})\boldsymbol{y}} \tag{B 2 5}$$

which is equivalent to

$$R^2 = 1 - \frac{\boldsymbol{e}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{e}}{\boldsymbol{y}'(\boldsymbol{S}^{-1} \otimes \boldsymbol{P})\boldsymbol{y}} \tag{B 2 6}$$

Some of the properties of McElroy's R^2 are as follows (see also Neudecker(1991))

- Clearly $0 \leq R^2 \leq 1$
- It is invariant under changes of location and changes of scale of $\boldsymbol{y}$
- It is related monotonically to an appropriate test statistic of $H_o: \overline{\boldsymbol{\gamma}} = \boldsymbol{0}$ against $H_a: \overline{\boldsymbol{\gamma}} \neq \boldsymbol{0}$
- It may be interpreted as the proportion of the sample variation in $(\boldsymbol{A} \otimes \boldsymbol{I}_n)\boldsymbol{y}$ explained by $(\boldsymbol{A} \otimes \boldsymbol{I}_n)\boldsymbol{U}$

The last result given above shows the principal weakness of McElroy's R^2 that is, the interpretation has only asymptotic justification because $\boldsymbol{S} = (\boldsymbol{A}'\boldsymbol{A})^{-1}$ is an asymptotic estimator of $\boldsymbol{\Sigma}$. Also McElroy's R^2 is derived on the assumption that each equation in the system includes an intercept. If this is not the case, then the decomposition in (B.2.4) which ensures the equivalence of the two formulas in (B.2.5) and (B.2.6) no longer applies. Finally, McElroy's R^2 measure is one for the goodness-of-fit of the entire seemingly unrelated regressions model. If the usual single-response coefficients of determination are calculated on the basis of the generalized least squares estimates and residuals, the usual decomposition of the ordinary least squares sums of squares will not hold in general, and ambiguous measures will be obtained. It is not clear what meaning can be attached to single-response goodness-of-fit measures in the context of seemingly unrelated regression models.

B.3 Optimization Algorithm by Price

Conlon (1992) presented a FORTRAN function, OPTCRS, which globally minimizes a constrained function using the Controlled Random Search (CRS) procedure of Price (1977). This procedure does not require that the function has derivatives and is well-suited for functions with local minima, that is, a minimum exists in a given (local) region of interest. The function OPTCRS concerns the following minimization problem.

$$\text{Minimize} \quad f(\theta) \quad \text{subject to} \quad \theta \in \Theta$$

where θ is a m-dimensional vector and Θ is bounded within a m-dimensional rectangle of finite volume in R^m.

Conlon (1992) describes the procedure as follows. The controlled random search algorithm involves storing function values (f_1 ... f_l) at l points distributed randomly throughout Θ. Let f_{max} and f_{min} be the largest and smallest function values in the storage. Let θ_{max} and θ_{min} be the corresponding points in the parameter space. At

each iteration a new point θ^* is generated by reflecting a randomly selected point from the storage through the centroid of m other points selected from the storage without replacement. If $f(\theta^*) < f_{max}$ then θ^* replaces θ_{max} in the storage. The process is stopped when f_{max} and f_{min} are close together.

The steps of algorithm are given below.

1. Load storage of size l by generating points $\theta_1, \ldots, \theta_l$ from Θ and also storing $f(\theta_1), \ldots, f(\theta_l)$.

2. For each iteration

   (a) generate a new point $\theta' \in \Theta$ as

   $$\theta' \leftarrow 2 \times \theta - \theta^r$$

   where θ^r is chosen randomly from $\theta_1, \ldots, \theta_l$ and θ is the centroid of m points chosen without replacement from $\theta_1, \ldots, \theta_l$.

   (b) Check $\theta' \in \Theta$. If not, repeat generation of θ'.

   (c) If $f(\theta') < f_{max}$ then replace f_{max} with $f(\theta')$, replace θ_{max} with θ' and determine new values for f_{max} and f_{min}.

   (d) If $f_{max} - f_{min} \leq$ TOL or the number of iterations has exceeded a cut-off then stop.

3. Return f_{min} as the minimum value and θ_{min} as the location of the minimum.

Conlon (1992) wrote the program using FORTRAN 77. It consists of the files *driver.f*, *optcrs.f*, and *optdb.f* and these are available from him upon request. The user needs only to modify the FORTRAN function *optfnc* in *driver.f* to suit his/her own needs. To run the program in the Department of Statistics computer system, the user must first obtain all the needed files and then compile the program by typing: *f77 -o driver driver.f optcrs.f optdb.f*. Then to invoke the program, just type: *driver*

For the first example given in Chapter 3 Section 3 4 the following code for *driver f* is used

```
      double precision optcrs,optfnc,f_x,xlim,wk,tol
      logical optbnd
      external optfnc,optbnd
      dimension x(3),xlim(2,3) wk(182)
      integer*4 s1 s2
      common /useeds/s1,s2
c
      call optsnd(641,4375)
      xlim(1,1)=-1 682d0
      xlim(1,2)=-1 682d0
      xlim(1,3)=-1 682d0
      xlim(2 1)=1 682d0
      xlim(2,2)=1 682d0
      xlim(2,3)=1 682d0
      nx=3
      np=30
      tol=1 00-3
      irmax=100
      itmax=1000
      f=optcrs(optfnc,optbnd,x,nx,xlim,np,irmax,itmax,tol,wk,ier)
      write(6,10) f,itmax,ier, x(1), x(2), x(3)
 10   format(' Function value = ' f10 6,' ie ' i5 ' iterations ' /,
     *   ' ier = ',i3, ' located at ', f10 6, f10 6, f10 6)
      END
c
```

```
      logical fuactioe optbnd(x,nx,xlim)
      double precision x(1),xlim(2,1), radsq
      date radsq/3 0/
      optbnd= TRUE
      do 10 i=1,nx
        if(x(i) lt xlim(1,i)  or  x(i) gt xlim(2,i)) optbnd= false
0     coetioue
      radius = x(1)*x(1) + x(2)*x(2) + x(3) * x(3)
      if (radius gt radsq) optbnd= false
      return
      end

      double precieion fnactiee optfoc(x,ns)
      double precieion x(1)  eighat
      double precieion beta0,beta1,beta2,beta3,beta11,beta22,beta33
      dimeneion eighat(2,2)
      data beta0/-1 860099294/,beta1/0 097382815/,beta2/0 000897626/,
     * beta3/-0 009052406/ beta11/0 102848418/,beta22/-0 001424141/,
     * beta33/-0 003191473/
      date eighat(1,1)/1/,eighat(1,2)/0/,eighat(2,1)/1/,eighat(2,2)/0/
      optfoc= beta0 + beta1*x(1) + beta2*x(2) + beta3*x(3)
     *    + beta11*x(1)*x(1) + beta22*x(2)*x(2) + beta33*x(3)*x(3)
      return
      end
```

APPENDIX C

DESIGN-EXPERT RESULTS FOR CHAPTER 3, SECTION 3.4

Numeric Optimization, File = EXAMPLE1 Run on 07/28/96 at 16 44 17

RES	RESPONSE	UNITS	OBS	MINIMUM	MAXIMUM	TRANS	MODEL
Y1	sprogiesse		17	1 370	1 920	None	Red Quad
Y2	thiacid es		17	20 16	51 59	None	Red Quad
Y3	cook loss		17	16 80	30 60	None	Red Quad
Y4	white indx		17	29 31	52 70	None	Red Lin

Numerical Optimization Results Cycle = 1

FACTOR	MIN	MAX	START	FINISH
x1	-1 000	1 000	0 6470	-0 6201
x2	-1 000	1 000	-0 3952	1 0000
x3	-1 000	1 000	-4 94E-02	0 9958

OPTIMIZATION PARAMETERS WEIGHTS

RESPONSE	SCALE	GOAL	LOW	HIGH	1ST	2ND	RESULT	d

springiness	Orig	max	1 3	2 5	1 00	1 920	0 52
thioacid no	Orig	min	17	51	1 00	19 13	0 94
cook loss	Orig	min	14	30	1 00	17 57	0 78
white indx	Orig	max	29	60	1 00	48 98	0 64

Desirability Function = 0 7018

Numerical Optimization Results Cycle = 2

FACTOR	MIN	MAX	START	FINISH
x1	-1 000	1 000	-0 9490	-0 6191
x2	-1 000	1 000	-0 1455	1 0000
x3	-1 000	1 000	-0 9246	0 8959

			OPTIMIZATION PARAMETERS			WEIGHTS		
RESPONSE	SCALE	GOAL	LOW	HIGH	1ST	2ND	RESULT	d
springiness	Orig	max	1 3	2 5	1 00		1 920	0 52
thioacid no	Orig	min	17	51	1 00		19 13	0 94
cook loss	Orig	min	14	30	1 00		17 57	0 78
white indx	Orig	max	29	60	1 00		48 98	0 64

Desirability Function = 0 7018

Numerical Optimization Results Cycle = 3

FACTOR	MIN	MAX	START	FINISH
x1	-1 000	1 000	-0 7186	-0 6192

x2	-1 000	1 000	-0 4718
x3	-1 000	1 000	-0 9658

RESPONSE	SCALE	GOAL	LOW	HIGH	1ST 2ND	RESULT	d
springiness	Orig	max	1 3	2 5	1 00	1 920	0 52
thiocid co	Orig	min	17	51	1 00	19 13	0 94
cook loss	Orig	min	14	30	1 00	17 57	0 78
white indx	Orig	max	29	60	1 00	48 98	0 64

Desirability Function = 0 7018

Numerical Optimization Results Cycle = 4

FACTOR	MIN	MAX	START	FINISH
x1	-1 000	1 000	-0 8878	-0 8205
x2	-1 000	1 000	0 2991	1 0000
x3	-1 000	1 000	-0 9954	0 9995

RESPONSE	SCALE	GOAL	LOW	HIGH	1ST 2ND	RESULT	d
springiness	Orig	max	1 3	2 5	1 00	1 920	0 52
thiocid co	Orig	min	17	51	1 00	19 13	0 94
cook loss	Orig	min	14	30	1 00	17 57	0 78
white indx	Orig	max	29	60	1 00	48 98	0 64

Desirability Function = 0 7018

Numerical Optimization Results Cycle = 5

FACTOR	MIN	MAX	START	FINISH
x1	-1 000	1 000	0 9370	-0 6196
x2	-1 000	1 000	0 3811	1 0000
x3	-1 000	1 000	-0 2082	0 9998

OPTIMIZATION PARAMETERS WEIGHTS

RESPONSE	SCALE	GOAL	LOW	HIGH	1ST	2ND	RESULT	d
springiness	Orig	max	1 3	2 5	1 00		1 920	0 52
thiocid ac	Orig	min	17	51	1 00		19 13	0 94
cook loss	Orig	min	14	30	1 00		17 57	0 78
white indx	Orig	max	29	60	1 00		48 98	0 64

Desirability Function = 0 7016

Numerical Optimization Results Cycle = 6

FACTOR	MIN	MAX	START	FINISH
x1	-1 000	1 000	0 7066	-0 6031
x2	-1 000	1 000	-0 3699	1 0000
x3	-1 000	1 000	-0 2670	0 9999

OPTIMIZATION PARAMETERS WEIGHTS

RESPONSE	SCALE	GOAL	LOW	HIGH	1ST	2ND	RESULT	d
springiness	Orig	max	1 3	2 5	1 00		1 921	0 52

thioacid co	Orig	min	17	51	1 00	19 11	0 94
cook loss	Orig	min	14	30	1 00	17 58	0 78
white indx	Orig	max	28	60	1 00	48 98	0 64

Desirability Function = 0 7018

Numerical Optimization Results Cycle = 7

FACTOR	MIN	MAX	START	FINISH
x1	-1 000	1 000	0 5411	-0 6145
x2	-1 000	1 000	0 6809	1 0000
x3	-1 000	1 000	-0 4632	0 8995

			OPTIMIZATION PARAMETERS			WEIGHTS	
RESPONSE	SCALE	GOAL	LOW	HIGH	1ST 2ND	RESULT	d
springiness	Orig	max	1 3	2 5	1 00	1 921	0 52
thioacid co	Orig	min	17	51	1 00	19 12	0 94
cook loss	Orig	min	14	30	1 00	17 58	0 78
white indx	Orig	max	28	60	1 00	48 98	0 64

Desirability Function = 0 7018

Numerical Optimization Results Cycle = 8

FACTOR	MIN	MAX	START	FINISH
x1	-1 000	1 000	-9 51E-02	-0 6180
x2	-1 000	1 000	-0 7853	1 0000

x3	-1 000	1 000	0 5805	0 9998

RESPONSE	SCALE	GOAL	LOW	HIGH	1ST WEIGHT	2ND WEIGHT	RESULT	d
sprlegisese	Orig	max	1 3	2 5	1 00		1 920	0 52
thioacid co	Orig	min	17	51	1 00		19 13	0 94
cook loss	Orig	min	14	30	1 00		17 57	0 78
white indx	Orig	max	29	60	1 00		48 98	0 64

Desirability Function = 0 7018

Numerical Optimization Results Cycle = 9

FACTOR	MIN	MAX	START	FINISH
x1	-1 000	1 000	0 2460	-0 5128
x2	-1 000	1 000	-0 9800	1 0000
x3	-1 000	1 000	0 1925	0 9998

RESPONSE	SCALE	GOAL	LOW	HIGH	1ST WEIGHT	2ND WEIGHT	RESULT	d
sprlegisese	Orig	max	1 3	2 5	1 00		1 921	0 52
thioacid co	Orig	min	17	51	1 00		19 12	0 94
cook loss	Orig	min	14	30	1 00		17 58	0 78
white indx	Orig	max	29	60	1 00		48 98	0 64

Desirability Function = 0 7018

Numerical Optimization Results Cycle = 10

FACTOR	MIN	MAX	START	FINISH
x1	-1 000	1 000	0 4216	-0 6180
x2	-1 000	1 000	-0 8301	1 0000
x3	-1 000	1 000	0 2761	0 9999

			OPTIMIZATION PARAMETERS		WEIGHTS		
RESPONSE	SCALE	GOAL	LOW	HIGH	1ST	2ND RESULT	d
springiness	Orig	max	1 3	2 5	1 00	1 920	0 52
thioacid ns	Orig	min	17	51	1 00	19 13	0 94
cook loss	Orig	min	14	30	1 00	17 57	0 78
whits indx	Orig	max	29	60	1 00	48 98	0 64

Desirability Function = 0 7016

APPENDIX D

PROOFS OF THEOREMS 4.5.1 AND 4.5.2

Before we prove Theorems 4.5.1 and 4.5.2 we first recall some of the notations used in Chapter 4. We have $\boldsymbol{X}^* = (\boldsymbol{M} \quad \boldsymbol{X})$, $\boldsymbol{M} = \boldsymbol{I}_r \otimes 1_n$, $\tilde{\boldsymbol{Z}} = diag(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \quad , \boldsymbol{Z}_r)$, $\tilde{\boldsymbol{Z}}_* = (\boldsymbol{I}_n - \frac{1}{n}\boldsymbol{J}_n)\tilde{\boldsymbol{Z}}$, $\boldsymbol{\mathcal{J}} = \boldsymbol{I}_r \otimes \boldsymbol{J}_n$, $\boldsymbol{F} = diag(\boldsymbol{f}_1(\boldsymbol{x}), \boldsymbol{f}_2(\boldsymbol{x}), \quad \boldsymbol{f}_r(\boldsymbol{x}))$, and $\boldsymbol{G} = diag(\boldsymbol{g}_1(\boldsymbol{x}), \boldsymbol{g}_2(\boldsymbol{x}) \quad \boldsymbol{g}_r(\boldsymbol{x}))$ with $\boldsymbol{g}_i'(\boldsymbol{x}) = [\boldsymbol{f}_i'(\boldsymbol{x}) \quad \frac{1}{b_i}\boldsymbol{1}_{b_i}']$, $i = 1, 2, \quad r$.

D.1 Proof of Theorem 4.5.1

Consider model (4.5.10) that is $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \tilde{\boldsymbol{Z}}\boldsymbol{\tau} + \boldsymbol{\epsilon}$. Under condition (4.5.5) of orthogonal blocking and by applying the result given in (4.5.9), namely,

$$\boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\tilde{\boldsymbol{Z}} = \boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\frac{1}{n}\boldsymbol{\mathcal{J}}\tilde{\boldsymbol{Z}} \tag{D.1.1}$$

the generalized least squares estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\tau}}$ satisfy the normal equations

$$\boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\frac{1}{n}\boldsymbol{\mathcal{J}}\tilde{\boldsymbol{Z}}\hat{\boldsymbol{\tau}} = \boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} \tag{D.1.2}$$

and

$$\tilde{\boldsymbol{Z}}'\frac{1}{n}\boldsymbol{\mathcal{J}}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}\hat{\boldsymbol{\beta}} + \tilde{\boldsymbol{Z}}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\tilde{\boldsymbol{Z}}\hat{\boldsymbol{\tau}} = \tilde{\boldsymbol{Z}}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} \tag{D.1.3}$$

It is clear that $\boldsymbol{M}\boldsymbol{M}' = \boldsymbol{\mathcal{J}}$. Using this result we can now rewrite (D.1.2) as

$$\boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{M}(\frac{1}{n}\boldsymbol{M}'\tilde{\boldsymbol{Z}}\hat{\boldsymbol{\tau}}) + \boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} \tag{D.1.4}$$

If we now premultiply equation (D 1 3) by $diag(\mathbf{1}'_{b_1}\ \mathbf{1}'_{b_2}\ \ldots\ \mathbf{1}'_{b_r})$ we get

$$\boldsymbol{M}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_t)\boldsymbol{Z}\boldsymbol{\tau} + \boldsymbol{M}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_t)\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{M}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} \qquad \text{(D 1 5)}$$

since $diag(\mathbf{1}'_{b_1}\ \mathbf{1}'_{b_2}\ \ldots\ \mathbf{1}'_{b_r})\boldsymbol{Z}' = \boldsymbol{M}'$ and $diag(\mathbf{1}'_{b_1}, \mathbf{1}'_{b_2}, \ldots, \mathbf{1}'_{b_r})\boldsymbol{Z}'\frac{\mathbf{1}}{n}\boldsymbol{\mathcal{J}}' = \boldsymbol{M}'$. Using the result

$$\boldsymbol{M}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z} = \boldsymbol{M}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\frac{\mathbf{1}}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z},$$

we can rewrite equation (D 1 5) in the form

$$\boldsymbol{M}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\frac{\mathbf{1}}{n}\boldsymbol{\mathcal{J}}\boldsymbol{Z}\boldsymbol{\tau} + \boldsymbol{M}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{M}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} \qquad \text{(D 1 6)}$$

If we substitute $\boldsymbol{M}\boldsymbol{M}'$ for $\boldsymbol{\mathcal{J}}$ we can express equation (D 1 6) as

$$\boldsymbol{M}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{M}(\frac{\mathbf{1}}{n}\boldsymbol{M}'\boldsymbol{Z}\boldsymbol{\tau}) + \boldsymbol{M}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{M}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} \qquad \text{(D 1 7)}$$

Next we look at the case where $\boldsymbol{\delta} = \boldsymbol{0}$. If $\boldsymbol{\delta} = \boldsymbol{0}$ then the generalized least squares estimators of $\boldsymbol{\beta}'_0$ and $\boldsymbol{\beta}$ satisfy the normal equations

$$\boldsymbol{M}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{M}\boldsymbol{\beta}_0 + \boldsymbol{M}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{M}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y}, \qquad \text{(D 1 8)}$$

and

$$\boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{M}\boldsymbol{\beta}_0 + \boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} \qquad \text{(D 1 9)}$$

By comparing equation (D 1 7) with equation (D 1 8) and equation (D 1 4) with equation (D 1 9) it follows that $\boldsymbol{\beta}$, which satisfies equations (D 1 4) and (D 1 7), is the same as $\boldsymbol{\beta}$ which satisfies equations (D 1 8) and (D 1 9).

D.2 Proof of Theorem 4.5.2

To prove Theorem 4.5.2 we need to consider the following lemmas

Lemma D.2.1 *Let the matrices $\boldsymbol{X}^*$, $\boldsymbol{M}$ and $\boldsymbol{F}$ be defined as before, then*

$$\frac{1}{n}\boldsymbol{M}'\boldsymbol{Z}\hat{\boldsymbol{\tau}} + \boldsymbol{F}'\boldsymbol{\beta} = \boldsymbol{H}\,[\boldsymbol{X}^{*\prime}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}^*]^{-1}\boldsymbol{X}^{*\prime}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} \tag{D.2.1}$$

Proof. From the normal equations (D.1.4), (D.1.7), (D.1.8) and (D.1.9) given in the proof of Theorem 4.5.1, we get

$$\boldsymbol{M}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)(\boldsymbol{M} \quad \boldsymbol{X})\begin{pmatrix} \tilde{\boldsymbol{\beta}}_0 \\ \boldsymbol{\beta} \end{pmatrix} = \boldsymbol{M}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)(\boldsymbol{M} \quad \boldsymbol{X})\begin{pmatrix} \frac{1}{n}\boldsymbol{M}'\boldsymbol{Z}\boldsymbol{\tau} \\ \boldsymbol{\beta} \end{pmatrix}$$

and

$$\boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)(\boldsymbol{M} \quad \boldsymbol{X})\begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta} \end{pmatrix} = \boldsymbol{X}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)(\boldsymbol{M} \quad \boldsymbol{X})\begin{pmatrix} \frac{1}{n}\boldsymbol{M}'\boldsymbol{Z}\boldsymbol{\tau} \\ \boldsymbol{\beta} \end{pmatrix}$$

The above results imply that

$$\begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\boldsymbol{M}'\boldsymbol{Z}\boldsymbol{\tau} \\ \boldsymbol{\beta} \end{pmatrix},$$

since the matrix $[\boldsymbol{M} \quad \boldsymbol{X}]$ is of full column rank. Thus if we premultiply the two sides by $\boldsymbol{H}' = [\boldsymbol{I}_r \quad \boldsymbol{F}']$, we get

$$\begin{pmatrix} \beta_{01} + \boldsymbol{f}_1'(\boldsymbol{x})\boldsymbol{\beta}_1 \\ \tilde{\beta}_{02} + \boldsymbol{f}_2'(\boldsymbol{x})\boldsymbol{\beta}_2 \\ \\ \beta_{0r} + \boldsymbol{f}_r'(\boldsymbol{x})\boldsymbol{\beta}_r \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\boldsymbol{1}_n'\boldsymbol{Z}_0\boldsymbol{\tau}_1 + \boldsymbol{f}_1'(\boldsymbol{x})\boldsymbol{\beta}_1 \\ \frac{1}{n}\boldsymbol{1}_n'\boldsymbol{Z}_0\boldsymbol{\tau}_2 + \boldsymbol{f}_2'(\boldsymbol{x})\boldsymbol{\beta}_2 \\ \\ \frac{1}{n}\boldsymbol{1}_n'\boldsymbol{Z}_0\tilde{\boldsymbol{\tau}}_r + \boldsymbol{f}_r'(\boldsymbol{x})\boldsymbol{\beta}_r \end{pmatrix} \tag{D.2.2}$$

Note that the left-hand side is equal to

$$\boldsymbol{H}'\begin{pmatrix} \tilde{\beta}_{01} \\ \beta_{02} \\ \\ \beta_{0r} \\ \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \\ \boldsymbol{\beta}_r \end{pmatrix} = \boldsymbol{H}\,[\boldsymbol{X}^{*\prime}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}^*]^{-1}\boldsymbol{X}^{*\prime}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y}$$

based on equation (4 5 12) and the right-hand side is equal to

$$\begin{pmatrix} \frac{1}{n}\mathbf{1}_n'\boldsymbol{Z}_0\boldsymbol{\tau}_1 + \boldsymbol{f}_1'(\boldsymbol{x})\bar{\boldsymbol{\beta}}_1 \\ \frac{1}{n}\mathbf{1}_n'\boldsymbol{Z}_0\boldsymbol{\tau}_2 + \boldsymbol{f}_2'(\boldsymbol{x})\bar{\boldsymbol{\beta}}_2 \\ \\ \frac{1}{n}\mathbf{1}_n'\boldsymbol{Z}_0\boldsymbol{\tau}_r + \boldsymbol{f}_r'(\boldsymbol{x})\bar{\boldsymbol{\beta}}_r \end{pmatrix} = \boldsymbol{I}_r[\frac{1}{n}\boldsymbol{M}'\boldsymbol{Z}\boldsymbol{\tau}] + \boldsymbol{F}'\boldsymbol{\beta}$$

It follows that

$$\frac{1}{n}\boldsymbol{M}'\boldsymbol{Z}\boldsymbol{\tau} + \boldsymbol{F}'\boldsymbol{\beta} = \boldsymbol{H}'[\boldsymbol{X}^{w'}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}^w]^{-1}\boldsymbol{X}^{w'}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y}$$

Lemma D.2.2 Let

$$\boldsymbol{T}' = \frac{1}{b}\boldsymbol{\mathcal{E}}' - \frac{1}{n}\boldsymbol{M}'\boldsymbol{Z} \tag{D 2 3}$$

where $\boldsymbol{\mathcal{E}} = \boldsymbol{I}_r \otimes \mathbf{1}_b$. Then

$$\boldsymbol{T}'\bar{\boldsymbol{\tau}} = \boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} \tag{D 2 4}$$

Proof Consider equation (D 1 3) If we premultiply both sides of this equation by $\boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}$ we get

$$\begin{aligned} \boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}\boldsymbol{Z}\frac{1}{n}\boldsymbol{\mathcal{J}}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{T}'\boldsymbol{\tau} \\ = \boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} \end{aligned}$$

Thus to prove equation (D 2 4) we only need to show that

$$\boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}\boldsymbol{Z}'\frac{1}{n}\boldsymbol{\mathcal{J}}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{O} \tag{D 2 5}$$

This can be accomplished by showing that

$$\boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}\boldsymbol{Z}'\boldsymbol{\mathcal{J}}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n) = \boldsymbol{O}$$

Let $\boldsymbol{\mathcal{F}}$ be defined as

$$\boldsymbol{\mathcal{F}} = (\boldsymbol{Z}_0'\boldsymbol{Z}_0)^{-1} = \mathrm{diag}(\frac{1}{n_1} \quad \frac{1}{n_2} \quad \cdots \quad \frac{1}{n_b}) \tag{D 2 6}$$

where n_j, $j = 1, 2, \ldots, b$, is the number of the jth block. Then $[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1} = [\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{Z}_0'\boldsymbol{Z}_0]^{-1} = \boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}}$. Next note that $\boldsymbol{Z}'\boldsymbol{\mathcal{J}}'$ is a block-diagonal matrix of order $br \times nr$ given by $\boldsymbol{Z}'\boldsymbol{\mathcal{J}}' = \boldsymbol{I}_r \otimes \boldsymbol{\mathcal{N}}$, where $\boldsymbol{\mathcal{N}}$ is the $b \times n$ matrix

$$\boldsymbol{\mathcal{N}} = \begin{pmatrix} n_1 & n_1 & & n_1 \\ n_2 & n_2 & & n_2 \\ & & & \\ n_b & n_b & & n_b \end{pmatrix}$$

Let

$$\boldsymbol{\mathcal{A}}_{ij} = \begin{pmatrix} \frac{n_{1i}}{n_1} & 0 & & 0 \\ 0 & \frac{n_{2i}}{n_2} & & 0 \\ & & & \\ 0 & 0 & & \frac{n_{bi}}{n_b} \end{pmatrix}, \quad i, j = 1, 2, \ldots, r \tag{D.2.7}$$

then $\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}} = (\boldsymbol{\mathcal{A}}_{ij})$. From these results we obtain $(\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}})\boldsymbol{Z}'\boldsymbol{\mathcal{J}}' = \boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{H}}$, where $\boldsymbol{\mathcal{H}}$ is a $b \times n$ matrix consisting of elements equal to one. It follows that

$$\boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}\boldsymbol{Z}'\boldsymbol{\mathcal{J}}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n) = \boldsymbol{T}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{H}})(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)$$

The right hand side of the above equation can also be written as

$$\boldsymbol{T}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{H}})(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n) = \boldsymbol{T}'(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\mathcal{H}}\boldsymbol{I}_n) = \boldsymbol{T}'(\boldsymbol{I}_r \otimes \boldsymbol{\mathcal{H}})$$

Using equation (D.2.3) we can rewrite $\boldsymbol{T}'(\boldsymbol{I}_r \otimes \boldsymbol{\mathcal{H}})$ as

$$\boldsymbol{T}'(\boldsymbol{I}_r \otimes \boldsymbol{\mathcal{H}}) = \frac{1}{b}\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_r \otimes \boldsymbol{\mathcal{H}}) - \frac{1}{n}\boldsymbol{M}'\boldsymbol{Z}(\boldsymbol{I}_r \otimes \boldsymbol{\mathcal{H}})$$

To prove equation (D.2.5) we only need to show that $\frac{1}{b}\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_r \otimes \boldsymbol{\mathcal{H}}) = \frac{1}{n}\boldsymbol{M}'\boldsymbol{Z}(\boldsymbol{I}_r \otimes \boldsymbol{\mathcal{H}})$. Now since $\frac{1}{b}\mathbf{1}_b'\boldsymbol{\mathcal{H}} = \mathbf{1}_n'$, then $\frac{1}{b}\boldsymbol{\mathcal{E}}'(\boldsymbol{I}_r \otimes \boldsymbol{\mathcal{H}}) = \boldsymbol{M}'$. Also, $\frac{1}{n}\boldsymbol{M}'\boldsymbol{Z}(\boldsymbol{I}_r \otimes \boldsymbol{\mathcal{H}}) = \boldsymbol{M}'$ since $\mathbf{1}_n'\boldsymbol{Z}_0 = (n_1, n_2, \ldots, n_b)$ and $\frac{1}{n}(n_1, n_2, \ldots, n_b)\boldsymbol{\mathcal{H}} = \mathbf{1}_n'$. From these results, we obtain

$$\boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}\boldsymbol{Z}'\boldsymbol{\mathcal{J}}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n) = \boldsymbol{O} \tag{D.2.8}$$

and hence

$$\boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}\boldsymbol{Z}'\frac{1}{n}\boldsymbol{\mathcal{J}}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{O}$$

Lemma D.2.1 The covariance between the quantities $\boldsymbol{H}'[\boldsymbol{X}^{s'}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}^s]^{-1}\boldsymbol{X}^{s'}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y}$ *and* $\boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y}$ *is zero, that is*

$$cov\{\boldsymbol{H}'[\boldsymbol{X}^{s'}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}^s]^{-1}\boldsymbol{X}^{s'}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y}\ \boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y}\} = \boldsymbol{O} \tag{D 2 9}$$

Proof To prove equation (D 2 9) it is sufficient to show that

$$\boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}^s = \boldsymbol{O} \tag{D 2 10}$$

Let $\boldsymbol{N} = (n_1\ n_2 \quad n_s)'$ Using the result

$$\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{M} = \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{Z}_0'\boldsymbol{1}_n = \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{N}$$

we get

$$\boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}\boldsymbol{Z}\,(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{M} = \boldsymbol{T}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}})(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{N})$$

Now since $\boldsymbol{\mathcal{F}}\boldsymbol{N} = \boldsymbol{1}_s$ we can simplify $\boldsymbol{T}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}})(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{N})$ to the form $\boldsymbol{T}'(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\mathcal{F}}\boldsymbol{N}) = \boldsymbol{T}'(\boldsymbol{I}_r \otimes \boldsymbol{1}_s) = \boldsymbol{T}'\boldsymbol{\mathcal{E}}$, with $\boldsymbol{T}'\boldsymbol{\mathcal{E}} = \frac{1}{s}\boldsymbol{\mathcal{E}}'\boldsymbol{\mathcal{E}} - \frac{1}{n}\boldsymbol{M}'\boldsymbol{Z}\boldsymbol{\mathcal{E}}$ where $\boldsymbol{\mathcal{E}}$ is defined in Lemma D 2 2 Both $\frac{1}{s}\boldsymbol{\mathcal{E}}'\boldsymbol{\mathcal{E}}$ and $\frac{1}{n}\boldsymbol{M}'\boldsymbol{Z}\boldsymbol{\mathcal{E}}$ are equal to $\boldsymbol{I}_r$ so that $\boldsymbol{T}'\boldsymbol{\mathcal{E}} = \boldsymbol{O}$ that is

$$\boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{M} = \boldsymbol{O} \tag{D 2 11}$$

If we postmultiply the two sides of equation (D 2.8) by $\boldsymbol{X}$ and use equation(D 1 1) we also obtain

$$\boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X} = \boldsymbol{O} \tag{D 2 12}$$

Equation (D 2 10) follows directly from equations (D 2 11) and (D 2 12)

Lemma D.2.4 Recall that $\boldsymbol{\mathcal{F}} \quad diag(\frac{1}{n_1}\ \frac{1}{n_2}, \quad \frac{1}{n_s})$ *and* $\boldsymbol{\mathcal{E}} = \boldsymbol{I}_r \otimes \boldsymbol{1}_s$ *Then*

1 $\boldsymbol{\mathcal{E}}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}})\boldsymbol{\mathcal{E}} = [\sum_{i=1}^{s} \frac{1}{n_i}]\boldsymbol{\Sigma}$

2. $\boldsymbol{\mathcal{E}}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}})\boldsymbol{Z}'\boldsymbol{M} = b\boldsymbol{\Sigma}$ and

3. $\boldsymbol{M}'\boldsymbol{Z}(\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}})\boldsymbol{Z}'\boldsymbol{M} = n\boldsymbol{\Sigma}$

Proof Recall that $\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}} = (\boldsymbol{\mathcal{A}}_{ij})$ where $\boldsymbol{\mathcal{A}}_{ij}$ is defined in equation (D 2 7) Note that $\boldsymbol{Z}'\boldsymbol{M} = \boldsymbol{I}_r \otimes \boldsymbol{N}$ and $\boldsymbol{N} = (n_1 \; n_2 \; \ldots \; n_b)$

1. Since $\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}} = (\boldsymbol{\mathcal{A}}_{ij})$ and $\boldsymbol{1}_b'\boldsymbol{\mathcal{A}}_{ij}\boldsymbol{1}_b = \sigma_{ij}\sum_{l=1}^{b}\frac{1}{n_l}$ then $\boldsymbol{\mathcal{E}}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}})\boldsymbol{\mathcal{E}} = [\sum_{l=1}^{b}\frac{1}{n_l}]\boldsymbol{\Sigma}$

2. $\boldsymbol{\mathcal{E}}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}})\boldsymbol{Z}'\boldsymbol{M} = \boldsymbol{\mathcal{E}}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}})(\boldsymbol{I}_r \otimes \boldsymbol{N}) = \boldsymbol{\mathcal{E}}'(\boldsymbol{\Sigma} \otimes \boldsymbol{1}_b) = b\boldsymbol{\Sigma}$

3. $\boldsymbol{M}'\boldsymbol{Z}(\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}})\boldsymbol{Z}'\boldsymbol{M} = \boldsymbol{M}'\boldsymbol{Z}(\boldsymbol{\Sigma} \otimes \boldsymbol{1}_b) = n\boldsymbol{\Sigma}$

Proof of Theorem 4.5.2.

Let us write $\boldsymbol{\eta}(\boldsymbol{x}) = \boldsymbol{G}'\boldsymbol{\theta}$ as

$$\boldsymbol{\eta}(\boldsymbol{x}) = \boldsymbol{G}'\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\tau}_1 \\ \boldsymbol{\beta}_2 \\ \boldsymbol{\tau}_2 \\ \\ \boldsymbol{\beta}_r \\ \boldsymbol{\tau}_r \end{pmatrix} = \begin{pmatrix} \boldsymbol{f}_1'(\boldsymbol{x})\boldsymbol{\beta}_1 + \frac{1}{b}\boldsymbol{1}_b'\boldsymbol{\tau}_1 \\ \boldsymbol{f}_2'(\boldsymbol{x})\boldsymbol{\beta}_2 + \frac{1}{b}\boldsymbol{1}_b'\boldsymbol{\tau}_2 \\ \\ \boldsymbol{f}_r'(\boldsymbol{x})\boldsymbol{\beta}_r + \frac{1}{b}\boldsymbol{1}_b'\boldsymbol{\tau}_r \end{pmatrix}$$

The right-hand side of the above equation in turn equals $\boldsymbol{F}'\boldsymbol{\beta} + \frac{1}{b}\boldsymbol{\mathcal{E}}'\boldsymbol{\tau}$ so that

$$\boldsymbol{\eta}(\boldsymbol{x}) = \boldsymbol{F}'\boldsymbol{\beta} + \frac{1}{b}\boldsymbol{\mathcal{E}}'\boldsymbol{\tau} \tag{D 2 13}$$

We apply the result given in Lemma D 2 1 to equation (D 2 13) and obtain

$$\boldsymbol{\eta}(\boldsymbol{x}) = \boldsymbol{H}'[\boldsymbol{X}^{o'}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}^{o}]^{-1}\boldsymbol{X}^{o'}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} + [\frac{1}{b}\boldsymbol{\mathcal{E}}' - \frac{1}{n}\boldsymbol{M}'\boldsymbol{Z}]\hat{\boldsymbol{\tau}} \tag{D 2 14}$$

From Lemma D 2 2 we can rewrite equation (D 2 14) as

$$\boldsymbol{\eta}(\boldsymbol{x}) = \boldsymbol{H}'[\boldsymbol{X}^{o'}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}^{o}]^{-1}\boldsymbol{X}^{o'}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} + \boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y} \tag{D 2 15}$$

Using equation (D 2 15) and Lemma D 2 4 the variance-covariance matrix of η is

$$\begin{aligned} Var[\hat{\eta}(\boldsymbol{x})] &= Var[\boldsymbol{H}'[\boldsymbol{X}^{s'}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_{\perp})\boldsymbol{X}^{s}]^{-1}\boldsymbol{X}^{s'}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y}] \\ &\quad + Var[\boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y}] \\ &= \boldsymbol{H}'[\boldsymbol{X}^{s'}(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{X}^{s}]^{-1}\boldsymbol{H} + \boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}\boldsymbol{T} \end{aligned}$$

which is

$$Var[\boldsymbol{\eta}(\boldsymbol{x})] = Var[\boldsymbol{\eta}_0(\boldsymbol{x})] + \boldsymbol{T}'[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}]^{-1}\boldsymbol{T} \qquad \text{(D 2 16)}$$

We can rewrite equation (D 2 16) in the following form

$$Var[\boldsymbol{\eta}(\boldsymbol{x})] = Var[\boldsymbol{\eta}_0(\boldsymbol{x})] + \boldsymbol{T}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}})\boldsymbol{T}$$

by using equation (D 2 6). Consider the second term on the right-hand side of the above equation. From equation (D 2 3) given in Lemma D 2 2 we have

$$\begin{aligned} &\boldsymbol{T}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}})\boldsymbol{T} = \\ &\tfrac{1}{b^2}\boldsymbol{\mathcal{E}}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}})\boldsymbol{\mathcal{E}} - \tfrac{2}{bn}\boldsymbol{\mathcal{E}}'(\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}})\boldsymbol{Z}'\boldsymbol{M} + \tfrac{1}{n^2}\boldsymbol{M}'\boldsymbol{Z}(\boldsymbol{\Sigma} \otimes \boldsymbol{\mathcal{F}})\boldsymbol{Z}'\boldsymbol{M} \end{aligned} \qquad \text{(D 2 17)}$$

Finally we obtain the desired result that is

$$Var[\boldsymbol{\eta}(\boldsymbol{x})] = Var[\boldsymbol{\eta}_0(\boldsymbol{x})] + \left[\left(\frac{1}{b^2}\sum_{i=1}^{b}\frac{1}{n_i}\right) - \frac{1}{n}\right]\boldsymbol{\Sigma}$$

by applying the results given in Lemma D 2 4 to equation (D 2 17)

APPENDIX E

SAS RESULTS FOR CHAPTER 4, SECTION 4 6

The SAS System

N = 30 Regression Models for Dependent Variable Y1

Number in Model	R-square	Adjusted R-square	C(p)	Variables in Model
2	0 83953834	0 82847202	40 46952	X1 X2

NOTE The above variables are included in all models to follow

3	0 93063544	0 92320352	4 73352	X1X2
3	0 84544524	0 82888580	40 02266	X2SQ
3	0 84023035	0 82311217	42 18288	X1SQ
4	0 93654233	0 92714120	4 28666	X2SQ X1X2

4	0.9313274S	0.92115373	8.44687	X1SQ X1X2
4	0.84#1372S	0.82334276	41.73598	X1SQ X2SQ
5	0.93723434	0.#2518402	6.00000	X1SQ X2SQ X1X2 ***

The SAS System 2

N = 32 Regression Models for Dependent Variable Y2

Number in Model	R-square	Adjusted R-square	C(p)	Variables in Model
2	0.76214196	0.74573798	27.89794	X1 X2

NOTE: The above variables are included in all models to follow

3	0.85492086	0.83937668	8.69152	X1X2
3	0.78304869	0.75980390	24.88691	X2SQ
3	0.77083103	0.74638792	27.61748	X1SQ
4	0.87582780	0.85743159	[illegible] 98049	X2SQ X1X2
4	0.86370994	0.84351883	8.71103	X1SQ X1X2

1	0 79183775	0 76099890	24 90642	X1SQ X2SQ
5	0 88461867	0 86242756	6 00000	X1SQ X2SQ X1X2 ***

N = 32 Regression Models for Dependent Variable Y3

Number in Model	R-square	Adjusted R-square	C(p)	Variables in Model
2	0 82314406	0 81094710	8 84103	X1 X2

NOTE The above variables are included in all models to follow

3	0 84546482	0 83001462	6 24679	X1X2
3	0 83830162	0 82097880	7 85496	X1SQ
3	0 82954355	0 81128035	9 68032	X2SQ
4	0 88182238	0 84112200	5 26071	X1SQ X1X2
4	0 85286431	0 83106643	6 98607	X2SQ X1X2
4	0 84470111	0 82189386	8 59424	X1SQ X2SQ

4	0.85802187	0.84264146	6.00000	X1SQ X2SQ X1X2 ***

N = 32 Regression Models for Dependent Variable Y3SQRT

Number in Model	R-square	Adjusted R-square	C(p)	Variables in Model
2	0.87457274	0.86592259	15.97748	X1 X2

NOTE The above variables are included in all models to follow

3	0.90073792	0.89010269	9.22083	X2SQ
3	0.88833754	0.87637371	13.37074	X1SQ
3	0.88238284	0.86978111	15.36380	X1X2
4	0.91450272	0.90183645	6.61388	X1SQ X2SQ
4	0.90884811	0.89499968	8.50675	X2SQ X1X2
4	0.89614774	0.88076222	12.75585	X1SQ X1X2
5	0.92231291	0.90737308	8.00000	X1SQ X2SQ X1X2 ***

Note The best models are indicated by ***

APPENDIX F

ESTIMATES OF VARIANCE-COVARIANCE MATRICES FOR SECTION 5 8

The following are the estimates of the matrices $\mathbf{\Sigma}_{\delta}$, $\mathbf{\Sigma}_{\Delta}$ and $\mathbf{\Sigma}$ used in the example in Section 5 8

$$\mathbf{\Sigma}_{\delta} = \begin{pmatrix} 0\ 5654 & 0\ 2094 & 0\ 2714 \\ 0\ 2094 & 0\ 0000 & 0\ 0597 \\ 0\ 2714 & 0\ 0597 & 0\ 0000 \end{pmatrix},$$

$$\mathbf{\Sigma} = \begin{pmatrix} 8\ 2743 & 0\ 9060 & -0\ 9928 \\ 0\ 9060 & 1\ 8794 & -0\ 0037 \\ -0\ 9928 & -0\ 0037 & 1\ 5333 \end{pmatrix}$$

and

$$\mathbf{\Sigma}_{\Delta} = \begin{bmatrix} \mathbf{\Sigma}^{11} & \mathbf{\Sigma}^{12} & \mathbf{\Sigma}^{13} \\ \mathbf{\Sigma}^{21} & \mathbf{\Sigma}^{22} & \mathbf{\Sigma}^{23} \\ \mathbf{\Sigma}^{31} & \mathbf{\Sigma}^{32} & \mathbf{\Sigma}^{33} \end{bmatrix}$$

where

$$\mathbf{\Sigma}^{11} = diag\{0\ \ 0\ 2574\ \ 0\ \ 0\ \ 0\}$$

$$\mathbf{\Sigma}^{12} = \begin{pmatrix} 0\ 0827 & -0\ 0663 & 0\ 1165 & -0\ 1022 & -0\ 1238 \\ 0\ 1321 & -0\ 0193 & 0\ 0672 & -0\ 0233 & -0\ 0654 \\ 0\ 0525 & -0\ 0634 & -0\ 0691 & -0\ 0143 & 0\ 0138 \\ 0\ 0136 & 0\ 0877 & 0\ 0241 & -0\ 1062 & -0\ 0353 \\ 0\ 1282 & -0\ 0742 & 0\ 0714 & -0\ 0687 & -0\ 1049 \end{pmatrix}$$

$$\Sigma^{13} = \begin{pmatrix} 0\,0590 & -0\,0465 & 0.1250 & -0\,1280 & -0\,0468 \\ 0\,0543 & 0\,0756 & 0\,1125 & -0\,0849 & 0\,0263 \\ 0\,0176 & 0\,0210 & 0\,0441 & -0\,0928 & 0\,0192 \\ 0\,0109 & 0\,0004 & -0\,0542 & -0\,0091 & -0\,0011 \\ 0\,0263 & -0\,0102 & 0.0770 & -0\,0727 & 0\,0672 \end{pmatrix}$$

$$\Sigma^{21} = \begin{pmatrix} 0\,0827 & 0\,1321 & 0\,0525 & 0\,0136 & 0\,1282 \\ -0\,0663 & -0\,0193 & -0\,0634 & 0\,0677 & -0\,0742 \\ 0\,1165 & 0\,0672 & -0\,0691 & 0\,0241 & 0\,0714 \\ -0\,1022 & -0\,0233 & -0\,0143 & -0\,1062 & -0\,0687 \\ -0\,1238 & -0\,0654 & 0\,0138 & -0\,0353 & -0\,1049 \end{pmatrix}$$

$$\Sigma^{22} = diag(0\,1373, 0\,0184\ 0\ 0\ 0),$$

$$\Sigma^{23} = \begin{pmatrix} 0\,0986 & 0\,0321 & 0\,0978 & -0\,0773 & 0\,0846 \\ -0\,0588 & -0\,0214 & -0\,0447 & 0\,0369 & -0\,0559 \\ 0\,0205 & -0\,0023 & 0\,0445 & -0\,0565 & 0\,0068 \\ -0\,0262 & 0\,0015 & -0\,0568 & 0\,0308 & -0\,0152 \\ 0\,0237 & 0\,0269 & -0\,0328 & 0\,0384 & 0\,0406 \end{pmatrix}$$

$$\Sigma^{31} = \begin{pmatrix} 0\,0590 & 0\,0543 & 0\,0176 & 0\,0109 & 0\,0263 \\ -0\,0465 & 0\,0756 & 0\,0210 & 0\,0004 & -0\,0102 \\ 0\,1250 & 0\,1125 & 0\,0441 & -0\,0542 & 0\,0770 \\ -0\,1280 & -0\,0849 & -0\,0928 & -0\,0091 & -0\,0727 \\ -0\,0468 & 0\,0263 & 0\,0192 & -0\,0011 & 0\,0672 \end{pmatrix}$$

$$\Sigma^{32} = \begin{pmatrix} 0\,0986 & -0\,0588 & 0\,0205 & -0\,0262 & 0\,0237 \\ 0\,0321 & -0\,0214 & -0\,0023 & 0\,0015 & 0\,0269 \\ 0\,0978 & -0\,0447 & 0\,0445 & -0\,0568 & -0\,0328 \\ -0\,0773 & 0\,0369 & -0\,0565 & 0\,0308 & 0\,0384 \\ 0\,0846 & -0\,0559 & 0\,0068 & -0\,0152 & 0\,0406 \end{pmatrix},$$

$$\Sigma^{33} = diag(0, 0\ 0\ 0\ 0)$$

REFERENCES

Adhikary B and Panda, R (1990) Sequential Method of Conducting Rotatable Designs. *Sankhya–B* **52** 212–218

Biles W E. (1975) A Response Surface Method for Experimental Optimization of Multiresponse Processes *Indust. Eng Chem. Proc Des and Develop* 15 152–158

Box, G E. P and Draper N R. (1959) A Basis for the Selection of a Response Surface Design *Journal of the American Statistical Association*, **54**, 622–654

Box G E P and Draper N R. (1965) The Bayesian Estimation of Common Parameters from Several Responses *Biometrika*, **52**, 355–365

Box G E P and Draper N R. (1975) Robust Designs *Biometrika*, **62** 347–352

Box G E P and Draper N R (1987) *Empirical Model–Building and Response Surfaces* John Wiley New York.

Box G E. P and Hunter J S (1957) Multifactor Experimental Designs for Exploring Response Surfaces *Annals of Mathematical Statistics* **28** 195–241

Box, G E P and Hunter J S (1961a) The 2^{k-p} Fractional Factorial Designs Part I *Technometrics* **3** 311–351

Box G E P and Hunter J S (1961b) The 2^{k-p} Fractional Factorial Designs Part II *Technometrics* **3** 449–458

Box G E P Hunter W G MacGregor J F and Erjavec J (1973) Some Problems Associated with the Analysis of Multiresponse Data. *Technometrics* 15 33–51

Box, G E P and Wilson K B (1951) On the Experimental Attainment of Optimum Conditions *Journal of the Royal Statistical Society* **B13** 1–45

Chen P M Varga D M Mielke E A and Drake S R. (1989) Poststorage Behavior of Apple Fruit After Low Oxygen Storage as Influenced by Temperature During Storage and in Transit *Journal of Food Science*, **54** 993–996

Conlon M (1992) The Controlled Random Search Procedure for Function Optimization *Communications in Statistics - Simulation* **21(3)** 919–923

Conlon M and Khuri A I (1988) MR: Multiple Response Optimization. *University of Florida - Department of Statistics - Technical Report 322*

Cook R and Nachtsheim C (1989) Computer-aided Blocking of Factorial and Response-surface Designs *Technometrics* **31** 339–346

Del Castillo E. and Montgomery, D C (1993) A Nonlinear Programming Solution to the Dual Response Problem *Journal of Quality Technology* **25** 199–204

Derringer G and Suich R (1980) Simultaneous Optimization of Several Response Variables *Journal of Quality Technology* **12** 214–219

Design-Expert Software Version 4.0 User's Guide (1993) Stat-Ease Incorporated Minneapolis

Dey, A and Das, M. N (1970) On Blocking Second Order Rotatable Designs. *Calcutta Statistical Association Bulletin*, **19** 75–85

Dodge M Kinata C and Stinson C (1995) *Running Microsoft Excel for Windows 95* Microsoft Press Washington

Fedorov V V (1972) *Theory of Optimal Experiments*, Academic Press, New York.

Harrington E C Jr (1965) The Desirability Function *Industrial Quality Control*, **12** 494–498

Jones E R. and Mitchell, T J (1978) Design Criteria for Detecting Model Inadequacy *Biometrika*, **65** 541–551

Kang G and Bates D M (1990) Approximate Inferences in Multiresponse Regression Analysis *Biometrika*, **77** 321–331

Karson M J Manson A R and Hader J R (1969) Minimum Bias Estimation and Experimental Design for Response Surfaces *Technometrics* **11** 461–475

Khuri A I (1985) A Test for Lack of Fit of a Linear Multiresponse Model *Technometrics*, **27** 213–218

Khuri A I (1990a) The Effect of Response Scaling on the Detection of Linear Dependencies Among Multiresponse Data. *Metrika*, **37** 217–231

Khuri A I (1990b) Multiresponse Rotatab[illegible] *Journal of Statistical Planning and Inference* **25** 1–6

Khuri A I (1992) Response Surface Models with Random Block Effects *Technometrics* **34** 26–37

Khuri A I (1994) Effect of Blocking on the Estimation of a Response Surface *Journal of Applied Statistics* **21** 305–316.

Khuri A I (1996a) Multiresponse Surface Methodology In *Handbook of Statistics, Volume 13-C R Rao and S Ghosh (eds) (in press)* North Holland Publishing Co Amsterdam.

Khuri A I (1996b) Response Surface Models with Mixed Effects *Journal of Quality Technology* **28** 177–186

Khuri A I and Conlon M (1981) Simultaneous Optimization of Multiple Responses Represented by Polynomial Regression Functions. *Technometrics* **23** 363–375

Khuri, A I and Cornell, J A (1996) *Response Surfaces*, 2d ed Marcel Dekker New York.

Kiefer, J (1972) Optimum Designs for Fitting Biased Multiresponse Surfaces *Proceedings of the Third International Symposium on Multivariate Analysis*, 287 297

Kim W and Draper N R. (1994) Choosing a Design for Straight Line Fits to Two Correlated Responses. *Statistica Sinica*, **4** 275–280

McElroy M B (1977) Goodness of Fit Test for Seemingly Unrelated Regressions *Journal of Econometrics* **6** 381–387

Myers, R. H (1971) *Response Surface Methodology* Allyn and Bacon Boston.

Myers, R. H and Carter, W H Jr (1973) Response Surface Techniques for Dual Response Systems *Technometrics* **15** 301 317

Myers R H Khuri A I and Carter W H Jr (1989) Response Surface Methodology 1966–1988 *Technometrics* **31** 137–157

Myers R. H and Montgomery, D C (1995) *Response Surface Methodology* John Wiley New York

Neudecker H and Windmeijer F A G (1991) R^2 in Seemingly Unrelated Regression Equations *Statistica Neerlandica*, **45** 405–411

Price W L (1977) A Controlled Random Search Procedure for Global Optimisation. *The Computer Journal*, **20** 367 370

Richert S H Morr C V and Cooney C M (1974) Effect of Heat and Other Factors Upon Foaming Properties of Whey Protein Concentrates *Journal of Food Science* **39**

42–48

Searle S R. (1982) *Matrix Algebra Useful for Statistics* John Wiley, New York

Seber G A. F (1984) *Multivariate Observations* John Wiley New York.

Sparks R. S. (1987) Selecting Estimators and Variables in the Seemingly Unrelated Regression Model *Communications in Statistics Simulation* **16**(1) 99–127

Srivastava, V K. and Giles D E. A (1987) *Seemingly Unrelated Regression Equations Models* Marcel Dekker New York

Tsao C L, Deng, J C., Cornell, J A Khuri A I and Schmidt, R H (1983) Effect of Washing Treatment on Quality of Minced Mullet Flesh *Journal of Food Science*, **48** 163–167

Wijesinha, M and Khuri A. I (1987a) The Sequential Generation of Multiresponse D-optimal Designs When the Variance-Covariance Matrix is Not Known *Communications in Statistics-Simulation*, **16** 239–259

Wijesinha, M and Khuri A I (1987b) Construction of Optimal Designs to Increase the Power of the Multiresponse Lack of Fit Test *Journal of Statistical Planning and Inference*, **16** 179–192.

Zellner A. (1962) An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. *Journal of the American Statistical Association* **57**, 348–368

BIOGRAPHICAL SKETCH

Elsie S. Valeroso was born in Quezon City, Philippines, on February 4, 1966, to Ely and Inocencia Valeroso. She received a Bachelor of Science degree in Statistics from the University of the Philippines, Diliman, Quezon City, in 1985. After two and a half years of teaching undergraduate mathematics and statistics at the University of the Philippines, she came to the United States to pursue graduate studies in Statistics. She obtained a Master of Science degree in Applied Statistics from Bowling Green State University, Bowling Green, Ohio in May 1989. She enrolled in the graduate program of the Department of Statistics, University of Florida, in Fall 1989. She currently teaches at Montana State University, Bozeman.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality as a dissertation for the degree of Doctor of Philosophy

André I. Khuri, Chairman
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy

John A. Cornell
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality as a dissertation for the degree of Doctor of Philosophy

Geoffrey Vining
Associate Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality as a dissertation for the degree of Doctor of Philosophy

Mang Tia
Professor of Civil Engineering

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy

August 1996

Dean, Graduate School